

Bowling Green State University
ScholarWorks@BGSU

Economics Faculty Publications

Economics

5-2008

Putting Expert Testimony in Its Epistemological Place: What Predictions of Dangerousness in Court Can Teach Us

M. Neil Browne

Bowling Green State University, nbrown2@bgsu.edu

Ronda R. Harrison-Spoerl

Follow this and additional works at: https://scholarworks.bgsu.edu/econ_pub

 Part of the [Law Commons](#)

Repository Citation

Browne, M. Neil and Harrison-Spoerl, Ronda R., "Putting Expert Testimony in Its Epistemological Place: What Predictions of Dangerousness in Court Can Teach Us" (2008). *Economics Faculty Publications*. 3.
https://scholarworks.bgsu.edu/econ_pub/3

This Article is brought to you for free and open access by the Economics at ScholarWorks@BGSU. It has been accepted for inclusion in Economics Faculty Publications by an authorized administrator of ScholarWorks@BGSU.

PUTTING EXPERT TESTIMONY IN ITS EPISTEMOLOGICAL PLACE: WHAT PREDICTIONS OF DANGEROUSNESS IN COURT CAN TEACH US

M. NEIL BROWNE*

RONDA R. HARRISON-SPOERL**

Judges and juries must make momentous and intricate decisions. The temptation is overwhelming for the court to request assistance from those who claim to know facts, interpretations, and explanatory models that promise to make those decisions more accurate.¹ As long as some of us know more than others about specific probative matters, courts will certainly seek to know what those experts know or, to anticipate, *what they claim to know*. But how can courts optimize their consumption of this expertise?

The use of expert knowledge to settle legal disputes is not a new phenomenon.² However, cases involving complex technical and scientific issues³ are becoming increasingly common, prompting what

* Distinguished Teaching Professor of Economics, Bowling Green State University.

** Psychologist, Independent Practice, Tallahassee, Florida.

1. The desire for expertise to provide sure resolutions to human dilemmas is aptly described in *Faust*. Johann Wolfgang von Goethe, *FAUST* 3–9 (Bayard Taylor trans., MacMillan Co. 1930) (1870) (The discussion in the Prologue among the poet, the Stage Manager and the Jester makes it clear that the poet has a more refined and nuanced appreciation of the kinds of questions theatre presents than the masses whom he demeans as he discusses the impending stage production.).

2. See Learned Hand, *Historical and Practical Considerations Regarding Expert Testimony*, 15 HARV. L. REV. 40, 41 n.2 (1901) (tracing the use of expert testimony to the thirteenth century); see also Steven R. Smith, *Mental Health Expert Witnesses: Of Science and Crystal Balls*, 7 BEHAV. SCI. & L. 145, 147 (1989).

3. See Ellen E. Deason, *Court-Appointed Expert Witnesses: Scientific Positivism Meets Bias and Deference*, 77 OR. L. REV. 59, 60 (1998) (“Technical, scientific, and complex subjects pose formidable challenges for the judicial system”); see also William W. Schwarzer, *Introduction to FEDERAL JUDICIAL CTR., REFERENCE MANUAL ON SCIENTIFIC EVIDENCE* 1, 1 (1994) (asserting the need for courts to comprehend issues that are increasingly “esoteric and complex”).

Although courts encounter complex scientific issues, many scholars suggest that the legal

some⁴ would call a proliferation of experts⁵ in the courtroom. More recently, judges have begun to recognize the potential for experts in social science research to assist the trier of fact⁶ in both civil and

system is not the best way to resolve scientific controversies. See, e.g., Paul Roberts, *Science in the Criminal Process*, 14 OXFORD J. LEGAL STUD. 469, 469 (1994) (noting the problematic relationship of science and law). For example, legal causation and scientific causation have very different standards. See, e.g., Matthew J. Smith, *The Admissibility of Expert Testimony and the Toxic Tort*, 15 J. PROD. & TOXICS LIAB. 97, 110 (1993) (describing the forms of evidence that the scientific community requires to prove causation as opposed to causation requirements in the legal standard of causation).

4. See James S. Laughlin, *When Students Confront the Experts: Toward Critical Thinking*, 81 ENG. J. 72, 72 (1992).

We surrender our car to the auto-repair expert, we surrender our children to the education specialist, we surrender our marriages to the sex therapists, our diet to the health experts, our beaches to the oil-spill experts, our defense to the military and foreign-policy experts in the Pentagon, our tax policy to the corporate-supported economic experts, [and] the administration of justice to the expert witnesses in the courtroom.

Id. See generally STEVEN BRINT, *IN AN AGE OF EXPERTS: THE CHANGING ROLE OF PROFESSIONALS IN POLITICS AND PUBLIC LIFE* (1994) (arguing, as the title suggests, that the role of professionals in politics and public life today are signs of a new “age of experts”).

5. See DOUGLAS WALTON, *APPEAL TO EXPERT OPINIONS* 1 (1997), (“[N]early everything we believe is believable because it is based on the opinion of experts. In this age of specialization and professionalization, it is not possible to escape accepting things on the basis of authority.”); see also Thomas L. Haskell, *Introduction to THE AUTHORITY OF EXPERTS: STUDIES IN HISTORY AND THEORY* xii (Thomas L. Haskell ed., 1984) (“[E]xperts have become so numerous and their knowledge and services have become so deeply interwoven with the fabric of our existence that some writers regard our reliance on expertise as the most distinctive feature of modern culture.”). James Laughlin agrees that “[a] dominant characteristic of this age is the proliferation of experts.” See Laughlin, *supra* note 4, at 72. Laughlin further argues, “Because we need those who can interpret this vast output of information and because an increasingly technological society demands a specialized workforce, the United States has come to rely on experts to guide the entire life of the nation.” *Id.* These needs have led to the increasing power of the expert in society. Magali Sarfatti Larson, *The Production of Expertise and the Constitution of Expert Power*, in *THE AUTHORITY OF EXPERTS: STUDIES IN HISTORY AND THEORY*, *supra*, at 28 (“Expertise . . . increasingly provides a base for attaining and exercising power by the people who can claim special knowledge in matters that their society considers important.”). See also Michael D. Bayles, *Professional Power and Self-Regulation*, 5 BUS. & PROF. ETHICS J. 26, 29 (1986) (arguing that much of the reason experts have power over non-experts is that in a majority of the situations in which an expert opinion is needed, non-experts are experiencing some type of a crisis). “The professional has the specialized knowledge and means to diagnose the problem, determine the alternative approaches to resolving it, and then take the necessary steps. This knowledge gives a professional power over an individual.” *Id.*

6. Perhaps the first recognition of social science research by courts was in *Muller v. Oregon*, 208 U.S. 412 (1908) and the “Brandeis Brief,” which was used to support the decision that regulation of the working hours of women was a legitimate governmental interest. See *id.* at 419 & n.†. Another notable early case, *Brown v. Board of Education*, 347 U.S. 483

criminal proceedings.⁷

Despite its frequent use,⁸ applying social science knowledge⁹ to legal questions entails persistent difficulties.¹⁰ Scholars have vigorously

(1954), and its acclaimed footnote eleven cited Kenneth Clark and other social psychologists' work on the detrimental effects of segregation. *Id.* at 494 n.11. See also NOREEN L. CHANNELS, *SOCIAL SCIENCE METHODS IN THE LEGAL PROCESS* 5 (1985), for a discussion of both *Muller* and *Brown*.

7. Experts have been called to testify on such widely diverse issues as eyewitness testimony, child abuse accommodation syndrome, post-traumatic stress disorder, and discrimination issues. See, e.g., *United States v. Coleman*, 41 M.J. 46, 47 (C.M.A. 1994) (involving an expert on child sex abuse); *United States v. Houser*, 36 M.J. 392, 394 (C.M.A. 1993) (describing rape trauma syndrome); *United States v. Suarez*, 35 M.J. 374, 375 (C.M.A. 1992) (involving descriptions of child abuse accommodation syndrome). See generally Saul M. Kassin et al., *The "General Acceptance" of Psychological Research on Eyewitness Testimony: A Survey of the Experts*, 44 AM. PSYCHOLOGIST 1089 (1989); Christopher Slobogin, *Psychiatric Evidence in Criminal Trials: To Junk or Not to Junk?*, 40 WM. & MARY L. REV. 1 (1998). Additionally, some experts have been called to testify as gang experts in gang injunction cases. See, e.g., Plácido G. Gómez, *It Is Not So Simply Because an Expert Says It Is So: The Reliability of Gang Expert Testimony Regarding Membership in Criminal Street Gangs: Pushing the Limits of Texas Rule of Evidence 702*, 34 ST. MARY'S L.J. 581, 605 (2003). Gómez argues that expert testimony to identify gang members relies heavily on experience and observation, and it is not reliable and thus should not be admissible. *Id.*

8. See Phoebe C. Ellsworth & Julius G. Getman, *Social Science in Legal Decision-Making*, in *LAW AND THE SOCIAL SCIENCES* 581, 581-88 (Leon Lipson & Stanton Wheeler eds., 1986) (discussing the frequency and variety of social science interactions with the law); Sharon D. Herzberger, *Social Science Contributions to the Law: Understanding and Predicting Behavior*, 25 CONN. L. REV. 1067, 1067 (1993) (describing the "substantial involvement of social scientists" in the legal arena).

Expert testimony may be used in various ways. See, e.g., L. Grant Foster, Comment, *A Case Study in Toxic Tort Causation: Scientific and Legal Standards Work Against Recovery for Victims*, 19 ENVTL. L. 141, 149 (1988) (explaining how expert testimony is often used to link an injury to exposure to a toxic substance). However, scholars often lament the use of expert testimony in the courtroom because their testimony often allows "junk science" into the courtroom. See, e.g., Ronald J. Allen & Joseph S. Miller, *The Common Law Theory of Experts: Deference or Education?*, 87 NW. U. L. REV. 1131, 1131 (1993) (noting the "increasingly controversial" nature of expert testimony).

9. See, e.g., Bruce D. Black, *The Use (or Abuse) of Expert Witnesses in Post-Daubert Employment Litigation*, 17 HOFSTRA LAB. & EMP. L.J. 269, 283 (2000). With regard to social sciences, Black says, "[T]he research, theories and opinions cannot have the exactness of hard science methodologies." *Id.* (quoting *Jensen v. Eveleth Taconite Co.*, 130 F.3d 1287, 1297 (8th Cir. 1997)).

10. See M. Neil Browne et al., *The Epistemological Role of Expert Witnesses and Toxic Torts*, 36 AM. BUS. L.J. 1, 39-40 (1998) (contrasting the legal system's methods for obtaining truth with truth that is sought by scientific processes); Sheila Jasanoff, *What Judges Should Know About the Sociology of Science*, 77 JUDICATURE 77, 80 (1993) ("[T]he ultimate goal of the courts is the attainable one of dispensing justice, not the impossible one of finding objective truth."); see also Peter W. Sperlich, *The Evidence on Evidence: Science and Law in Conflict and Cooperation*, in *THE PSYCHOLOGY OF EVIDENCE AND TRIAL PROCEDURE* 325, 337-342 (Saul M. Kassin & Lawrence S. Wrightsman eds., 1985) (describing jurists' concerns about scientific evidence); Thomas M. Crowley, *Help Me Mr. Wizard! Can We Really Have*

debated the extent to which courts should rely on mental health professionals' predictions of violence.¹¹ The introduction of these professionals' predictions of violence¹² has become a generally accepted practice, and yet, an examination of "dangerousness" research¹³ reveals

"Neutral" Rule 706 Experts?, 1998 DET. C.L. MICH. ST. U. L. REV. 927, 928-30 (asserting the danger of deluding oneself into thinking that any expert can be truly neutral); Emily C. Lieberman, *Forced Medication and the Need to Protect the Rights of the Mentally Ill Criminal Defendant*, 5 CARDOZO PUB. L. POL'Y & ETHICS J. 479, 480 (2007) (discussing the difficulty of predicting dangerousness).

11. See, e.g., JOHN W. PARRY, NATIONAL BENCHMARK ON PSYCHIATRIC AND PSYCHOLOGICAL EVIDENCE AND TESTIMONY 20 (1998); Charles P. Ewing, *"Dr. Death" and the Case for an Ethical Ban on Psychiatric and Psychological Predictions of Dangerousness in Capital Sentencing Proceedings*, 8 AM. J.L. & MED. 407, 409 (1983) ("Over the past two decades, empirical research has consistently demonstrated that psychiatric and psychological predictions of dangerousness generally prove to be inaccurate."); Eric S. Janus & Paul E. Meehl, *Assessing the Legal Standard for Predictions of Dangerousness in Sex Offender Commitment Proceedings*, 3 PSYCHOL. PUB. POL'Y & L. 33, 35 (1997) ("There is continuing debate about whether predictions of dangerousness are accurate enough to support deprivation of liberty."); Thomas Regnier, *Barefoot in Quicksand: The Future of "Future Dangerousness" Predictions in Death Penalty Sentencing in the World of Daubert and Kumho*, 37 AKRON L. REV. 469, 476-77 (2004); Gary Gleb, Comment, *Washington's Sexually Violent Predator Law: The Need to Bar Unreliable Psychiatric Predictions of Dangerousness from Civil Commitment Proceedings*, 39 UCLA L. REV. 213, 222-28 (1991) (describing psychiatric predictions of long-term dangerousness as "highly unreliable and prejudicial evidence"); Eugenia T. La Fontaine, Note, *A Dangerous Preoccupation with Future Danger: Why Expert Predictions of Future Dangerousness in Capital Cases Are Unconstitutional*, 44 B.C. L. REV. 207, 229-35 (2002) (arguing that expert testimony about future dangerousness is unreliable, inaccurate, and unconstitutional). But see Alexander D. Brooks, *The Constitutionality and Morality of Civilly Committing Violent Sexual Predators*, 15 U. PUGET SOUND L. REV. 709, 736-40 (1992) (suggesting that mental health professionals can accurately predict future sexual violence).

12. See, e.g., *Atkins v. Virginia*, 536 U.S. 304, 307-09, 321 (2002) (remanding a case in which a mentally retarded individual was sentenced to death for a capital crime, based in part on predictions of "future dangerousness"); *United States v. Wattleton*, 296 F.3d 1184, 1192-94 (11th Cir. 2002) (acquitting Defendant Wattleton on grounds of insanity, but committing him on grounds of dangerousness).

13. Many commentators have criticized the use of the word "dangerousness" as problematic. See John Monahan, *Introduction to THE CLINICAL PREDICTION OF VIOLENT BEHAVIOR: CRIME AND DELINQUENCY ISSUES* 4-5 (1981) ("'Dangerousness' confuses issues regarding *what* one is predicting with the *probability* one is assigning to its prediction."); Edwin I. Megargee, *The Prediction of Dangerous Behavior*, 3 CRIM. JUST. & BEHAV. 3, 5 (1976) ("'Dangerousness' is an unfortunate term, for it implies there is a trait of 'dangerousness' which, like intelligence, is a relatively constant characteristic of the person being assessed. However, the degree of danger an individual represents to himself or others varies markedly as a function of a number of variables."); see also Thomas Grisso & Paul S. Appelbaum, *Is It Unethical to Offer Predictions of Future Violence?*, 16 LAW & HUM. BEHAV. 621, 623 n.3 (1992).

Future discourse in this area might be facilitated by ridding ourselves of the phrase *predictions of dangerousness*. It has no logical meaning in the

the problems inherent in the field and in social science literature as a whole.

This Article challenges the widely held assumption that social science evidence generally, and predictions of violence specifically, are sufficiently objective, neutral, or “true” to warrant anything but a wary acceptance in the judicial system.¹⁴ Our argument will follow two paths. The first is the more abstract argument based on philosophy of science notions about what science is and what it can offer. When we examine the case law relevant to expert witnesses, we will see how the standards for admissibility of this testimony fail to adequately consider how scientific pronouncements are shaped¹⁵ by the “argumentative aspects of social life.”¹⁶ Science is more appropriately conceived of as a contextually based, social endeavor¹⁷ and is not capable of providing the objective, universal truths¹⁸ that justify presentation by a single expert.¹⁹

context of the behavioral and social sciences. To “predict” is to make a statement about the likelihood of a future event or behavior. *Dangerousness* seems to refer not to an event or behavior, but to a condition that exists as a function of the presence of someone or something perceived as “dangerous.”

Id. at 623 n.3. For the purposes of this portion of the Article, the terms “violent behavior” or “violence” will be used except when quoting sources.

14. The positivist perspective of science contends that the truth exists and is waiting for humans to find it. See DONALD POLKINGHORNE, *METHODOLOGY FOR THE HUMAN SCIENCES* 16–20 (1983) (describing the naturalism-empiricism-positivism tradition and its proposition that experience of the senses is the only source of knowledge); Richard Rorty, *Science as Solidarity*, in *THE RHETORIC OF THE HUMAN SCIENCES* 38, 38–39 (John S. Nelson et al. eds., 1987) (describing the notion that truth corresponds with reality); David C. Geary, *The Evolution of Cognition and the Social Construction of Knowledge*, 51 *AM. PSYCHOL.* 265, 266 (1996) (describing the basic assumption “that all knowledge is culturally mediated and is constructed through” social activities).

15. See Browne et al., *supra* note 10, at 72 (revealing that scientific knowledge is created in a social context that is full of power struggles).

16. MICHAEL BILLIG, *ARGUING AND THINKING: A RHETORICAL APPROACH TO SOCIAL PSYCHOLOGY* 4 (1987); see also Jeanine Czubaroff, *The Deliberative Character of Strategic Scientific Debates*, in *RHETORIC IN THE HUMAN SCIENCES* 28, 28–29 (Herbert W. Simons ed., 1989).

17. See SHEILA JASANOFF, *SCIENCE AT THE BAR: LAW, SCIENCE, AND TECHNOLOGY IN AMERICA* 52–53 (1995). “The authority of scientific claims derives, according to the sociological account, not directly, from the representation of physical reality, but indirectly, from the certification of claims through a multitude of informal, often invisible, negotiations among members of relevant disciplines.” *Id.* at 52.

18. See Bert Black et al., *Science and the Law in the Wake of Daubert: A New Search for Scientific Knowledge*, 72 *TEX. L. REV.* 715, 765 (1994). Black points out that law and science have different ends. He admonishes, “Lawyers and judges need to understand that the scientific landscape encompasses many hypotheses that lie between the poles of speculative conjecture and established scientific fact. There is no absolute level of certainty that makes a

The second component of this argument will demonstrate that, even accepting the logical empiricists' position that "truth" exists and can be revealed through scientific methodology, violence predictions have not attained the degree of objectivity, neutrality, or universalism frequently posited, and thus, need for presentation of multiple expert opinions exists in this particular illustration of social science expertise. Thus, Parts V and VI familiarize the reader with the empirical literature on the prediction of dangerousness and challenge some of the misconceptions about mental health professionals' ability to predict future violence.²⁰ We begin by providing a brief history of the "first-generation"²¹ dangerousness prediction literature and the legal community's response to this research. Part V includes a representative, although not comprehensive, review of recent methodological improvements in "second-generation"²² research, followed by an analysis of the current empirical evidence.

A key component in this discussion is the description and comparative analysis of clinical²³ and actuarial²⁴ prediction strategies

proposition scientific." *Id.*

Additionally, see Philip Mirowski, *The Rhetoric of Modern Economics*, 3 HIST. HUM. SCI. 243, 246 (1992), which recounts Descartes' inclination to "redefine rationality so as to isolate it from all emotional attachment[]." *Id.* By doing so he hoped to establish thought in the mind that would be as orderly and dependable as the natural order of numbers. *Id.* This yearning for certitude is both old and understandable, but yearning does not *ipso facto* create fulfillment of the desire.

19. Experts who suggest that experts are truth dispensers, rather than thoughtful people presenting an argument, are probably persuasive to uninformed listeners. *But see* ROBYN M. DAWES, *HOUSE OF CARDS: PSYCHOLOGY AND PSYCHOTHERAPY BUILT ON MYTH* viii (1994) (arguing that any mental health expert taking a position of certainty regarding the future behavior of a particular individual is "by definition incompetent" given the documented inability of such experts to justifiably make such a claim).

20. Inaccuracies in the literature concerning prediction accuracy abound. *See infra* notes 229–30 and accompanying text.

21. *See* John Monahan, *The Prediction of Violent Behavior: Toward a Second Generation of Theory and Policy*, 141 AM. J. PSYCHIATRY 10, 10 (1984). Professor Monahan coined this phrase—"first-generation"—to refer to the five seminal studies conducted in the early 1970s examining the accuracy of mental health professionals' predictions of violence. *Id.*; *see infra* notes 245–51 and accompanying text.

22. Monahan, *supra* note 21, at 10 (describing the methodological limitations of dangerousness research and calling for empirical research to address these methodological problems); *see infra* Part V.C.

23. Clinical prediction relies upon information gathered by an experienced clinician who then forms an estimation of risk intuitively or subjectively based on the information obtained. *See infra* Part VI.

24. Actuarial strategies rely on mathematical equations to predict violence based on the presence or absence of factors previously identified to be correlated with future violence. *See infra* notes 366–69. These equations produce numerical probability estimates describing the

presently used to inform dangerousness decision making. Prediction strategies have generally been divided into two types: clinical and actuarial.²⁵ Early prediction studies relied exclusively on clinical judgment while more recent research efforts have attempted to quantify prediction methods with the use of mathematical equations.

Critics challenging proffered predictions of dangerousness have based their objections on a few evidentiary doctrines including: the relevancy doctrine,²⁶ the reliability or “trustworthiness” standard²⁷ articulated in the Federal Rules of Evidence²⁸ and elaborated in *Daubert v. Merrell Dow Pharmaceuticals, Inc.*,²⁹ and the “unfair prejudice” doctrine.³⁰ The question of whether violence prediction testimony can

risk of future violent acts. *Id.* See generally *infra* Part VII.

25. This dichotomy is a vast over-generalization. Clinical judgments can be informed by numerical probability estimates, while actuarial instruments can incorporate subjective impressionistic data into the equation. See *infra* note 403 and accompanying text.

26. See *infra* notes 225–28 and accompanying text.

27. This requirement ensures that proffered testimony meets what the court considers as minimum standards of reliability and validity. Critics of opposing expert dangerousness predictions assert that existing data concerning violence prediction fails to demonstrate accurate, valid, or reliable findings to meet these standards. See, e.g., George E. Dix, *Expert Prediction Testimony in Capital Sentencing: Evidentiary and Constitutional Considerations*, 19 AM. CRIM. L. REV. 1, 21 (1981) (“Increasingly it is agreed that [clinical dangerousness] testimony, which is of dubious accuracy and questioned by the mental health professional community, is unlikely to aid the trier of fact to a significant degree.”); Ewing, *supra* note 11, at 409 (“[E]mpirical research has consistently demonstrated that psychiatric and psychological predictions of dangerousness generally prove to be inaccurate.”).

28. See FED. R. EVID. 702 (“If scientific, technical, or other specialized knowledge will assist the trier of fact to understand the evidence or to determine a fact in issue, a witness qualified as an expert by knowledge, skill, experience, training, or education may testify thereto in the form of an opinion or otherwise.”). See also Harvey Brown, *Eight Gates for Expert Witnesses*, 36 HOUS. L. REV. 743 (1999), for a comprehensive analysis of the evidentiary requirements of proffered expert testimony.

An addendum to Rule 702 was passed by the Committee on Rules of Practice and Procedure of the Judicial Conference of the United States in 2002 that requires “(1) the testimony is based upon sufficient facts or data, (2) the testimony is the product of reliable principles and methods, and (3) the witness has applied the principles and methods reliably to the facts of the case.” FED. R. EVID. 702. This committee is composed of twenty-five federal judges who determine the policy of United States courts. See 28 U.S.C. § 331 (2006).

29. 509 U.S. 579, 588 (1993) (asserting that the Federal Rules of Evidence superseded the *Frye v. United States*, 293 F. 1013 (D.C. Cir. 1923), “general acceptance” standard for evidence previously governing expert testimony); see also *Frye*, 293 F. at 1014. The Court concluded Rule 702 mandates a gatekeeping role for the judiciary. *Daubert*, 509 U.S. at 597. Judges performing this role are to ascertain whether proffered testimony provides a sufficiently reliable foundation based on scientific knowledge. *Id.* at 590. This has become known as the “evidentiary reliability” test. *Id.*

30. Courts are wary of the “aura of scientific infallibility” that may prejudice the trier of fact. See Margaret A. Berger, *Evidentiary Framework*, in REFERENCE MANUAL ON SCIENTIFIC EVIDENCE, *supra* note 3, at 37, 115. Both the Federal Rules of Evidence, FED. R.

withstand these challenges is examined and leads to the authors' conclusions that expert testimony based solely on clinical judgment should frequently fail to meet admissibility standards.

Therefore, we contend that the party providing this testimony must, on a case-by-case basis, bear the burden of proof in demonstrating the efficacy of violence prediction testimony by the proposed expert.³¹ Further, we acknowledge that empirical research supports the efficacy of actuarial based predictions,³² but conclude that no single, universally accepted actuarial instrument has been acknowledged at present.³³ Finally, the relative inefficacy of cross-examination to challenge unreliable or arbitrary prediction testimony requires presentation of opposing experts. Thus, we argue that the trier of fact must ascertain the appropriateness of the instrument in the particular case, as well as assess the competence of each proffered expert in using the assessment instrument before admitting proffered testimony of this type.

We use illustrative cases as a vehicle to explore the application of actuarial prediction testimony to statutory standards of proof requirements. Recommendations are made concerning the kinds of questions members of the judiciary should consider when evaluating proffered expert violence prediction testimony. Some general maxims are also provided concerning the types of questions and circumstances that might lead the trier of fact to determine whether the testimony fails to pass evidentiary muster.

I. SCIENCE MEETS LAW: THE ROLE OF EXPERTS IN THE COURTROOM

A. *Legal and Scientific Discourse*

Legal and scientific paradigms are separate and incongruous entities. The legal system is adversarial in nature:³⁴ each party in a dispute ideally

EVID. 403 ("Although relevant, evidence may be excluded if its probative value is substantially outweighed by the danger of unfair prejudice, confusion of the issues, or misleading the jury, or by considerations of undue delay, waste of time, or needless presentation of cumulative evidence."), and *Daubert*, 509 U.S. at 595, provide for the exclusion of relevant, but possibly prejudicial or misleading, evidence that may unduly sway the trier of fact. See Brown, *supra* note 28, at 750.

31. For the courts, efficacy must be demonstrated by empirical evidence admissible under the Federal Rules of Evidence. See *infra* Part II.

32. See *infra* notes 365–73 and accompanying text.

33. See VERNON L. QUINSEY ET AL., VIOLENT OFFENDERS: APPRAISING AND MANAGING RISK 169 (1998) (discussing the selection criteria for choosing one actuarial instrument over another).

34. See, e.g., DEBORAH L. RHODE, IN THE INTERESTS OF JUSTICE: REFORMING THE

has an equally skilled and knowledgeable advocate armed with the best arguments for his side,³⁵ and the jury arrives at the “just” conclusion by watching the battle between the conflicting arguments.³⁶

Lawyers on each side of a case present only that evidence favorable to their own clients.³⁷ The legal system relies on the cross-examination and refutation that takes place in the courtroom to clarify the evidence³⁸ and allow the jury to fill in the evidence gap on either side and make the best decision.³⁹ The “truth” that is sought by the legal process is thus not simply the most objective or descriptive conclusion, but a conclusion ensuring justice and fairness⁴⁰ instead of some elusive “truth.”⁴¹

LEGAL PROFESSION 56 (2000) (“The claim that adversarial clashes yield factually accurate results is not self-evident.”)

35. See, e.g., ARTHUR ISAK APPLBAUM, *ETHICS FOR ADVERSARIES: THE MORALITY OF ROLES IN PUBLIC AND PROFESSIONAL LIFE* 3–13 (1999).

36. The “best” arguments as determined by a jury are defined from a legal perspective, not an intellectual one. The adversarial principle is similar to the metaphor of the “market place of ideas.” If everyone presents his or her “wares” in the market, we will have the best opportunity to arrive at the “truth.” See Peter Huber, *Junk Science in the Courtroom*, 26 VAL. U. L. REV. 723, 733 (1992); Clifton T. Hutchinson & Danny S. Ashby, *Daubert v. Merrell Dow Pharmaceuticals, Inc.: Redefining the Bases for Admissibility of Expert Scientific Testimony*, 15 CARDOZO L. REV. 1875, 1878–80 (1994). This “market place of ideas” metaphor may not be appropriate for the courtroom, where there may not be equal availability of expertise, whether from lack of ability or financial resources, to the particular parties in the dispute. Such inequalities in representation would be not an equal exchange, but an oligopolistic situation, invalidating the competitive metaphor.

37. See *Confronting the New Challenges of Scientific Evidence*, 108 HARV. L. REV. 1481, 1587 (1995) [hereinafter *Confronting the New Challenges*].

38. Congress relied on this process when it enacted the Federal Rules of Evidence with its lenient standards of admissibility. See Leslie A. Lunnery, *Protecting Juries from Themselves: Restricting the Admission of Expert Testimony in Toxic Tort Cases*, 48 SMU L. REV. 103, 105 (1994).

39. *Confronting the New Challenges*, *supra* note 37, at 1587 (citing a case where, because lawyers failed to emphasize counterarguments, the jury believed that all asbestosis victims would have identical symptoms).

40. See, e.g., Michael S. Jacobs, *Testing the Assumptions Underlying the Debate About Scientific Evidence: A Closer Look at Juror “Incompetence” and Scientific “Objectivity,”* 25 CONN. L. REV. 1083, 1086 (1993).

41. See Browne et al., *supra* note 10, at 39 (recognizing that the legal system is in hot pursuit of justice, and discussing that, if truth were sought in the courtroom, juries would consist of the intellectual elite, not average members of the community); Jacobs, *supra* note 40, at 1086 (citing, e.g., Richard O. Lempert, *Civil Juries and Complex Cases: Let’s Not Rush to Judgment*, 80 MICH. L. REV. 68, 80–84 (1981)); ALEXIS DE TOCQUEVILLE, *DEMOCRACY IN AMERICA* 295 (Vintage Books 1945) (1840). De Tocqueville writes,

The jury, and more especially the civil jury, serves to communicate the spirit of the judges to the minds of all the citizens; and this spirit, with the habits which attend it, is the soundest preparation for free institutions. It imbues all classes with a respect for the thing judged and with the notion

In contrast, science is a pursuit for more knowledge and better understanding of the world, marked by openness and rational deliberation by its participants, under few or no time constraints.⁴² Objectivity and empirical observation are lauded as the prime determiners of scientific fact.⁴³

When an expert enters the courtroom, the two worlds of science and law are brought into play.⁴⁴ Despite the difference in approach and standards of proof for the two disciplines,⁴⁵ the specialized product

of right. If these two elements be removed, the love of independence becomes a mere destructive passion. It teaches men to practice equity; every man learns to judge his neighbor as he would himself be judged. And this is especially true of the jury in civil causes; for while the number of persons who have reason to apprehend a criminal prosecution is small, everyone is liable to have a lawsuit. . . . It invests each citizen with a kind of magistracy; it makes them all feel the duties which they are bound to discharge towards society and the part which they take in its government. By obliging men to turn their attention to affairs other than their own, it rubs off that private selfishness which is the rust of society.

DE TOCQUEVILLE, *supra*, at 295.

42. Science's advantage of lack of time pressure illustrates differences between the legal system and science. Unlike science, the legal system does face deadlines and short time periods. *But see Confronting the New Challenges*, *supra* note 37, at 1533 (suggesting that the distinction between the rigor required by law and science is a false dichotomy).

43. *See* Hutchinson & Ashby, *supra* note 36, at 1878–79. Although jurors may assume that scientists embody these attributes, other commentators reject the view of science as objective in any meaningful sense of that word; instead, a scientist in the courtroom is considered a parallel to the advocacy approach to truth used in our legal system. Stephen J. Gould is a prominent critic of claims of objectivity by scientists. *See, e.g.*, STEPHEN J. GOULD, *THE MISMEASURE OF MAN* 20 (1981); *see also* Daubert v. Merrell Dow Pharms., Inc., 509 U.S. 579, 590 (1993) (“[T]here are no certainties in science.”); Allan Megill, *Introduction: Four Senses of Objectivity*, in *RETHINKING OBJECTIVITY* 1, 5–8 (Allan Megill ed., 1994).

44. *See* Eileen A. Scallen & William E. Wiethoff, *The Ethos of Expert Witnesses: Confusing the Admissibility, Sufficiency and Credibility of Expert Testimony*, 49 HASTINGS L.J. 1143, 1143–44 (1998) (“[T]he testimonial discourse of experts, though not cast in the elegant form of oratory, has rhetorical tenor and effect. Expert testimony, even that based on natural or social science, is argumentation, made for, and in, a unique context—the law . . .”).

45. *See, e.g., Confronting the New Challenges*, *supra* note 37, at 1484. While considering the “inherent inconsistency” between the demands of the legal and scientific worlds, Dan Burk stated:

Lawyers, who dwell . . . in an adversarial realm, are governed by written rules of professional conduct that require them to . . . act as zealous advocates on their client's behalf. . . . Scientists, on the other hand . . . have developed strong unwritten professional rules—based on norms of intellectual objectivity—that are different from those governing lawyers.

offered by the expert⁴⁶—scientific knowledge⁴⁷—is in demand by the legal system to provide necessary input⁴⁸ on issues of scientific complexity.⁴⁹ As we seek an optimal role for experts in the legal system, it is essential to inquire into the origins and attributes of their expertise.

Dan L. Burk, *Using Scientists as Courtroom Witnesses: System Needs Improvement*, THE SCIENTIST, Oct. 26, 1992, at 11, 11.

Benjamin Cardozo, however, compared science to the legal system, saying, “The work of a judge is in one sense enduring and in another sense ephemeral. . . . In the endless process of testing and retesting, there is a constant rejection of the dross, and a constant retention of whatever is pure and sound and fine.” BENJAMIN CARDOZO, *THE NATURE OF THE JUDICIAL PROCESS* 178–79 (1921) (quoted in *Daubert*, 509 U.S. at 597 n.13).

The legal system’s time constraints are illustrated by its regulations on the number of experts who may testify and the time that may be given to hear their testimony. See, e.g., Jack B. Weinstein, *Scientific Evidence in Complex Litigation*, in TRIAL EVIDENCE, CIVIL PRACTICE, AND EFFECTIVE LITIGATION TECHNIQUES IN FEDERAL AND STATE COURTS 709, 726 (1991) (citing *Flannigan v. GAF Corp.*, 904 F.2d 36 (6th Cir. 1990)). The Court in *Daubert* also emphasized the legal system’s need to settle disputes “finally and quickly.” *Daubert*, 509 U.S. at 597.

Unfortunately, the structure of the adversary system discourages honest skepticism and reflection. See, e.g., Peter W. Huber, *On Law and Sciosophy*, 24 VAL. U. L. REV. 319, 347 (1990) (“Any half-competent lawyer will prefer committed support from the fringe to even the slightest ambivalence from the middle.”).

46. See, e.g., Robert Hairman, *The Rhetoric of Inquiry and the Professional Scholar*, in RHETORIC IN THE HUMAN SCIENCES, *supra* note 16, at 211, 224 (“Disciplinary knowledge is both a claim to know and a means of social control, . . . a body of knowledge and a suppression of interpretive thinking.”).

47. “Scientific knowledge,” according to *Daubert*, is evidence derived by the scientific method and supported by appropriate validation. 509 U.S. at 590. Evidence will assist the trier of fact if it is relevant. *Id.* at 591. This relevancy requirement however is not sufficient to meet the assistance criterion. *Id.* Instead, the evidence also must contain a “valid scientific connection to the pertinent inquiry.” *Id.* at 592.

48. In some states, this appeal to expertise is mandatory. See Stephen D. Easton, “Yer Outta Here!” *A Framework for Analyzing the Potential Exclusion of Expert Testimony Under the Federal Rules of Evidence*, 32 U. RICH. L. REV. 1, 8 (1998) (“In medical or other professional malpractice actions based upon the law of some states, plaintiffs must present admissible expert testimony about the standard of care and the defendant’s failure to meet this standard.”).

49. As confusing as this input may be, the inadequacy of our personal knowledge base to form adequate reasoning is highlighted by Haskell, *supra* note 5, at x, where he concedes that

If it could be done, I suppose all of us would prefer to base everything we do and think on “good reasons” rather than expert authority. But the conduct of everyday affairs requires us to hold so many opinions and make so many decisions that we cannot possibly base them on the personally examined evidence and the inwardly compelling logic that “good reasons” imply.

B. Expert Testimony: Uses and Misuses

Expert testimony in the courtroom is subject to certain pitfalls. A brief look at the roles of jury, judge, and expert when such testimony is heard gives a sense of the difficulty of evaluating expert testimony and the need for a clearer understanding of the nature and origins of expert testimony in what is essentially an epistemological issue.⁵⁰ The difficulty of sorting through expert testimony makes us wonder: Who should be responsible for this evaluation? Different admissibility standards⁵¹ produce different interpretative roles for the players.⁵²

High on the list of concerns associated with a court's use of expert testimony is misinterpretation of such testimony by the finder of fact.

Id.

50. By epistemology, we mean in this Article "How do we know what we think we know?" rather than "How do we know the *truth*?" The latter question is from a realist perspective we do not share, as will become plain.

51. However, as we shall see later, admissibility may not be the only, or even the prime concern in a reevaluation of experts. Tweaking admissibility standards may not address a more central issue: the origins of the evidence being admitted. *See* Scallen & Wiethoff, *supra* note 44, at 1149 ("[C]ourts have focused on the issue of the admissibility of expert testimony at the expense of understanding how issues of admissibility, sufficiency and credibility all work together to determine the ethos of expert testimony presented at trial."). The same could be said of many legal scholars who focus excessively on issues of admissibility. *Contra* Thomas D. Lyon & Jonathan J. Koehler, *The Relevance Ratio: Evaluating the Probative Value of Expert Testimony in Child Sexual Abuse Cases*, 82 CORNELL L. REV. 43, 43 (1996) (arguing that "[s]tandards pertaining to the admissibility of scientific testimony are critical to the outcome in many trials."). No doubt they are correct; however, the authors' suggestions for improving admissibility standards using a relevance ratio do not embrace the larger concerns of expert witness testimony. *See generally* Michael H. Graham, *The Expert Witness Predicament: Determining "Reliable" Under the Gatekeeping Test of Daubert, Kumho, and Proposed Amended Rule 702 of the Federal Rules of Evidence*, 54 U. MIAMI L. REV. 317 (2000). Graham points out significant ambiguities regarding the admissibility of expert witnesses' testimony in the courtroom, but he does not examine the consequences of expert testimony per se. *Id.* But *see generally* Bruce Abramson, *Blue Smoke or Science? The Challenge of Assessing Expertise Offered as Advocacy*, 22 WHITTIER L. REV. 723 (2001). Abramson's article is concerned with procedural issues aimed at keeping "bad science" out of the courtroom. Despite the lack of attention to experts' monopolization of the truth and the moral hazard surrounding their practice, he recognizes that "[e]xpert testimony is inherently problematic." *Id.* at 765; *see also* Lorie S. Gildea, *Sifting the Dross: Expert Witness Testimony in Minnesota After the Daubert Trilogy*, 26 WM. MITCHELL L. REV. 93, 93-94 (2000) (arguing that Minnesota should change its expert witness testimony admissibility criteria to that of *Daubert*).

52. The admissibility standards are based not only on a sense of roles, but of the legal goals of those players. *See* Weinstein, *supra* note 45, at 730. The use of experts varies depending on the legal goals; the goal of reducing litigation is served by a stringent standard for admission, while the goal of obtaining compensation for deserving plaintiffs may necessitate the admission of less pristine expertise. *See, e.g.,* Allen & Miller, *supra* note 8, at 1131 (noting the "increasingly controversial" nature of expert testimony).

Experts have been defined as follows:

[P]ersons who are qualified, either by actual experience or by careful study, so as to enable them to form definite opinions with respect to a division of science, branch of art, or department of trade about which persons having no particular training or special study are incapable of forming accurate opinions or of drawing correct conclusions.⁵³

This definition of expertise points to the usefulness of such specialized knowledge⁵⁴ as an aid to the jury in evaluating the evidence in many legal controversies.⁵⁵ However, the knowledge gap between expert and juror⁵⁶ that necessitates expert testimony leads to a different problem: If the role of the expert is to enable the juror to assess the evidence in the case, what happens when the juror is unable to assess the testimony of the expert?⁵⁷

53. 31A AM. JUR. 2D *Expert and Opinion Evidence* § 1 (2002) (citations omitted).

54. See Eric G. Jensen, Comment, *When "Hired Guns" Backfire: The Witness Immunity Doctrine and the Negligent Expert Witness*, 62 UMKC L. REV. 185, 186 (1993). Jensen explains, "With the increase in legislation and government regulation over the last thirty years, no one, not even a highly educated judge, can remain fully knowledgeable of all pertinent issues. Thus, the testimony of expert witnesses is often needed and used to clarify, explain and assist on many important issues." *Id.* at 189–90 (citations omitted).

55. The role of expert as educator for the jury has been the predominant view. See, e.g., Allen & Miller, *supra* note 8, at 1131–33, 1141 (noting that the Federal Rules of Evidence encourage the educational role of the expert).

56. *Id.* at 1133. This knowledge gap may be the largest in cases involving "novel scientific testimony." See *id.*

57. See Jacobs, *supra* note 40, at 1088 (stating "the oft-expressed notion that lay jurors 'are incompetent to evaluate scientific proof critically'" (quoting Edward J. Imwinkelried, *Judge Versus Jury: Who Should Decide Questions of Preliminary Facts Conditioning the Admissibility of Scientific Evidence?*, 25 WM. & MARY L. REV. 577, 580 (1984))). Given this inability to distinguish between "charlatans" and "Nobel prize winners," presented with two extremes, jurors are likely to presume the truth lies somewhere in between. *Id.* at 1088–89; see also Browne et al., *supra* note 10, at 13 n.67 ("[D]enunciation of juror competence and reliance on outside help to assess the reliability of the evidence should apply likewise to the jury's ability to evaluate the testimony that is admitted.").

Vidmar and Diamond, on the other hand, argue that juries are competent:

Jurors appear motivated to critically assess the content of the expert's testimony and weigh it in the context of the other trial evidence, as they are instructed to do. They appear to understand the nature of the adversary process, at least in the context of their specific trial. Even though many jurors may not have had prior exposure to the trial process, it appears that they develop an understanding from the give and take of

When the court hears the testimony of an “expert,” especially someone recognized as a “scientific expert,” the jury may be overly impressed⁵⁸ by the credentials presented⁵⁹ and terminology used by this individual,⁶⁰ hindering the jury’s ability to fully understand⁶¹ and

examination and cross-examination and exposure to opposing experts. Indeed, rather than simply deferring automatically to experts, as critics have claimed, the trial process appears to make them aware of the fallibility of expert testimony. This is not to say that every juror is motivated and grasps the expert testimony, because the data seldom shed light on the thought processes of individual jurors, but the deliberation process appears to result in closer examination of diverging views and understandings—just as the legal system assumes it does.

Neil Vidmar & Shari Seidman Diamond, *Juries and Expert Evidence*, 66 BROOK. L. REV. 1121, 1174 (2001). Vidmar and Diamond describe another study conducted by Vidmar involving five malpractice cases:

[These studies] show that the jurors were not passive in evaluating the experts or their testimony. Indeed, one of the findings from the interviews of jurors was that they clearly understood the adversary system. They identified basic disagreements between the experts. They considered absence of evidence and incompleteness of testimony. They scrutinized possible motives behind each expert’s testimony such as money and the possibility that an expert was giving testimony to support a fellow physician. They had a basically solid understanding of burdens of proof and where the expert testimony fit into assessing that burden. Most importantly, the jurors in each case evaluated the testimony in the context of other trial evidence.

Id. at 1141–42.

One recent proposal was to extend *Frye* and have the judge present the proffered scientific evidence to a committee of scientists who would testify at trial as to its validity. See Weinstein, *supra* note 45, at 728.

58. See PAULO FREIRE, PEDAGOGY OF THE OPPRESSED 58 (Myra Bergman Ramos trans., 1970) (stating that nonexperts are often led to act as “containers” or “receptacles to be filled” by unexamined pronouncements of expert knowledge).

59. See, e.g., John W. Osborne, *Judicial/Technical Assessment of Novel Scientific Evidence*, 1990 U. ILL. L. REV. 497, 501 (arguing that the jury may be more impressed by the string of credentials than by the actual testimony of the expert).

60. See, e.g., IAN R. FRECKELTON, THE TRIAL OF THE EXPERT: A STUDY OF EXPERT EVIDENCE AND FORENSIC EXPERTS 141 (1987) (describing how an expert can use technical or foreign language, uncommon words, and complex sentences); see also Stephen A. Saltzburg, *Improving the Quality of Jury Decisionmaking*, in VERDICT: ASSESSING THE CIVIL JURY SYSTEM 341, 363 (Robert E. Litan ed., 1993) (arguing that if a judge is unable to comprehend expert or scientific testimony, she should assume that the jury is similarly perplexed).

61. See Bayles, *supra* note 5, at 29 (“Because they allegedly lack the necessary knowledge, laypersons cannot even evaluate professionals’ recommendations or actions.”). The recognition of this paradox is, of course, nothing new. In his dialogue, *Charmides*, Plato has Socrates address this very issue. After questioning one of his students, Socrates eventually comes to conclude the following:

evaluate the evidence⁶² presented by the expert.⁶³ Commentators have noted the “aura of scientific infallibility”⁶⁴ that surrounds the expert,⁶⁵ especially when offering certain forms of evidence, such as statistical

Then assuredly, wisdom or temperance, if only a science of science, and of the absence of science or knowledge, will not be able to distinguish the physician who knows from one who does not know but pretends or thinks that he knows, or any other professor of anything at all; like any other artist, he will only know his fellow in art or wisdom, and no one else.

PLATO, *Charmides*, in THE DIALOGUES OF PLATO 22 (Benjamin Jowett trans., 1937). What Plato is arguing here is that those who are not physicians will never be able to distinguish between the physician who really knows what she is talking about and the physician who does not.

62. One expert in Texas, known as the “Doctor of Doom,” offered testimony in over seventy capital punishment cases that the defendant was a dangerous “sociopath who would kill again.” RHODE, *supra* note 34, at 104. Although this evidence was, in some cases, not based on personal examination of the defendant by the expert, juries imposed the death penalty in all but one trial. *Id.*

63. See Lunney, *supra* note 38, at 104–05; see also United States v. Downing, 753 F.2d 1224, 1236 (3d Cir. 1985) (recognizing that jurors may be swayed into believing that scientific testimony is especially reliable and trustworthy). But see United States v. Williams, 583 F.2d 1194, 1199 (2d Cir. 1978) (rejecting the defendant’s argument that the expert “awed [jurors] by an ‘aura of mystic infallibility’”); Weinstein, *supra* note 45, at 711 (noting that the fear that jurors will give too much weight to testimony from one labeled an “expert” has little evidence to support it); Neil Vidmar, *Pap and Circumstance: What Jury Verdict Statistics Can Tell Us About Jury Behavior and the Tort System*, 28 SUFFOLK U. L. REV. 1205, 1205–06 (1994) (arguing that statistics suggesting problems with jury decision making in tort cases are subject to numerous methodological problems). See generally Augustine Brannigan & Michael Lynch, *On Bearing False Witness: Credibility as an Interactional Accomplishment*, 16 J. CONTEMP. ETHNOGRAPHY 115 (1987).

64. See Weinstein, *supra* note 45, at 723.

65. One commentator suggests that this aura could be mitigated by a revelation of communications between experts and the attorneys for whom they testify. See Stephen D. Easton, *Ammunition for the Shoot-Out with the Hired Gun’s Hired Gun: A Proposal for Full Expert Witness Disclosure*, 32 ARIZ. ST. L.J. 465, 508–09 (2000). Easton wants to limit the partisanship of expert witnesses by making full disclosure of communication between expert witnesses and attorneys mandatory, and he wants to end attorney-crafting of expert witness testimony and give opposing lawyers assistance revealing expert witnesses’ biases. As Easton says,

In a system which holds itself out as one that is designed to search for the truth, one would expect that the substantial benefits enjoyed by experts should be accompanied by increased reporting about the formation of expert, as opposed to fact witness, testimony. Instead, as a practical matter, reporting regarding the formation of expert testimony is often more limited than discovery regarding the formulation of fact witness testimony.

Id. (citation omitted).

evidence.⁶⁶ When expert testimony from a single perspective is presented, the jury may be influenced to regard such testimony⁶⁷ as indisputably accurate⁶⁸ or decisive,⁶⁹ thus abandoning its role as decision maker.⁷⁰ When testimony from multiple perspectives is heard, the

66. See MORRIS E. CHAFETZ, *THE TYRANNY OF EXPERTS: BLOWING THE WHISTLE ON THE CULT OF EXPERTISE* 103 (1996), where he points out that the average American is particularly intimidated by mathematical and scientific knowledge. "A large part of the population, when asked about their proficiency in the mathematics, confessed that they were not 'number people.' They believe they cannot understand numbers and statistics, so they invest in the incantations of those who purport to unravel the mysteries of the universe: the scientists." *Id.*; see also WALTON, *supra* note 5, at 22 ("We tend to be intimidated by experts, not only because they so often use technical jargon but because we ourselves . . . are not in a good position to really understand the expert's reasons for advocating a particular conclusion or recommending a particular course of action.").

67. See IRVING M. COPI & CARL COHEN, *INTRODUCTION TO LOGIC* 165 (10th ed. 1998).

In attempting to make up one's mind about some difficult or complicated question, it is entirely reasonable to be guided by the judgment of an acknowledged expert who has studied the matter thoroughly. When we argue that a given conclusion is correct on the ground that an expert authority has come to that judgment, we commit no fallacy. Indeed, such recourse to authority is necessary for most of us on very many matters. Of course, an expert's judgment constitutes no conclusive proof[,] . . . but expert opinion surely is one reasonable way to support a conclusion.

Id.

68. *But see, e.g.*, Franklin Strier, *Making Jury Trials More Truthful*, 30 U.C. DAVIS L. REV. 95, 115 (1996) ("The adversary system is not a reliable means of bringing all the relevant scientific data to the adjudicator's attention or for separating valid research from unwarranted conclusions.") (citation omitted). Strier laments that because of the decision in *Daubert v. Merrell Dow Pharmaceuticals, Inc.*, 509 U.S. 579 (1993), "almost any practitioner's view, no matter how iconoclastic, may be welcome if reached via the scientific method." Strier, *supra*, at 115. He goes on to say that "[s]ocial scientists generally shirk the responsibility to expose the limits of their own expertise." *Id.*

69. See, e.g., Allan Raitz et al., *Determining Damages: The Influence of Expert Testimony on Jurors' Decision Making*, 14 LAW & HUM. BEHAV. 385 (1990). In a mock jury study, the researchers found that "many jurors wholeheartedly accepted the figures proffered by the expert for the plaintiff." *Id.* at 390. Specifically, the jurors awarded the exact amount mentioned by the plaintiff's expert: "[T]he analyses of monetary awards indicated that expert testimony influenced awards in an upward direction. Jurors unaided by expert testimony awarded significantly less than those exposed to an expert." *Id.* at 392; see also Weinstein, *supra* note 45, at 712 (noting the potential impact of the "imprimatur of the trial judge's decision that [the person testifying] is an 'expert'").

70. Some scholars have proposed educational measures for juries to help them better execute their role. See Keith Broyles, Note, *Taking the Courtroom into the Classroom: A Proposal for Educating the Lay Juror in Complex Litigation Cases*, 64 GEO. WASH. L. REV. 714, 745 (1996). Because "jurors do not enter the trial with a blank slate," and they must evaluate the testimony of expert witnesses, it is argued that they need assistance in their

resulting jury confusion⁷¹ can impel such coping tactics as “splitting the difference”⁷² between the two opinions.⁷³ Jurors may also be unduly

decision-making process. SAUL M. KASSIN & LAWRENCE S. WRIGHTSMAN, *THE AMERICAN JURY ON TRIAL: PSYCHOLOGICAL PERSPECTIVES* 86 (1988). For example, in one case, counsel furnished the jurors with definitions of all the exotic terms to be used in some patent litigation. See *CytoLogix Corp. v. Ventana Med. Sys., Inc.*, 424 F.3d 1168, 1170 (Fed. Cir. 2005). For example, Jack B. Weinstein has suggested the use of pedagogical aids to better allow the jury to organize the information. Weinstein, *supra* note 45, at 728; see also Stanley D. Davis, *A Fresh Look at Hypothetical Questions and Ultimate Issues: The Kansas Experience*, 36 U. KAN. L. REV. 311, 353 (1988). Jurors awed by the mental prowess of experts may not rely upon their own experience, even if they are instructed to do so. See Hutchinson & Ashby, *supra* note 36, at 1878.

71. This confusion may be heightened by the belief Chafetz points out: “Many people believe *someone* does, in fact, have answers that they do not possess. Because they want to believe in that magical ‘someone,’ they are defenseless against those who claim to have special knowledge.” CHAFETZ, *supra* note 66, at xiii.

[A]mong 170 million adult Americans, 27 million read below the fifth-grade level. Some 60 to 65 million read below the ninth grade level. To comprehend public policy discussions on the op-ed pages of the *New York Times*, the *Washington Post*, or the *Wall Street Journal*, a reader needs at least a twelfth-grade reading level. In other words, nearly two out of five Americans are ill-equipped to participate fully in public life. They do not have the resources available to them should they wish to question the scientists and would-be prognosticators.

Id. at 117–18.

Although courts and scholars may doubt the capabilities of the jury, protecting them from scientific “banter” between opposing experts may lead the jury to view the expert’s testimony as unquestionable. See Alan W. Tamarelli, Jr., *Daubert v. Merrell Dow Pharmaceuticals: Pushing the Limits of Scientific Reliability—The Questionable Wisdom of Abandoning the Peer Review Standard for Admitting Expert Testimony*, 47 VAND. L. REV. 1175, 1176 (1994) (arguing that other experts should make decisions about the reliability of scientific evidence).

72. In a situation where jurors are faced with testimony from two or more conflicting experts, each with his or her own “aura” of unquestionable accuracy and wisdom, they may try to compromise by “splitting the difference” between the multiple viewpoints. See Richard A. Epstein, *A New Regime for Expert Witnesses*, 26 VAL. U. L. REV. 757, 758 (1992); Jacobs, *supra* note 40, at 1090 (noting the concern that jurors may “split the intellectual difference” between the “charlatans and Nobel prize winners”); Allen & Miller, *supra* note 8, at 1132. The dichotomous thinking that presents scientific experts as either “charlatans” or “Nobel Prize winners” is also reflected in the label “junk science.” Such hasty categorization restricts the court’s understanding of expertise and its usefulness, and it substitutes for careful consideration of the merits of such testimony, particularly when “junk science” is used as a dismissal of the opponent’s testimony.

73. The response of jurors to conflicting opinions is described by Franklin Strier. *Supra* note 68, at 115. He explains that in the O.J. Simpson criminal trial, *People v. Simpson*, No. BA097211 (Cal. Super. Ct. L.A. County 1995), jurors became visibly disinterested as experts tepidly debated each other’s claims. Strier, *supra* note 68, at 115 n.65 (“What was supposed to be educative was instead combative and confounding.”).

influenced by the appearance⁷⁴ or speaking ability of the experts.⁷⁵ Furthermore, there are concerns that jurors might be persuaded by testimony from “hired guns”⁷⁶ or experts whose opinions are based on

74. See David L. Wiley, Comment, *Beauty and the Beast: Physical Appearance Discrimination in American Criminal Trials*, 27 ST. MARY'S L.J. 193, 234–35 (1995) (considering how jurors discriminate according to physical appearance and suggesting that some type of remedy should be made to ensure defendants are judged according to their actions instead of their appearance).

75. See FRECKELTON, *supra* note 60, at 146. Freckleton refers to a study in which an actor hired to play an expert giving a lecture was reviewed by three audiences of trained educators. These audiences awarded the lecture favorable reviews, even though it contained almost no substantive communication. FRECKELTON, *supra* note 60, at 146 (citing Donald H. Naftulin et al., *The Doctor Fox Lecture: A Paradigm of Educational Seduction*, 48 J. MED. EDUC. 630 (1973)). “This highlights the vital role that the presentation by the expert plays in his or her effectiveness in purveying a point of view to the tribunal of fact.” *Id.*

An expert's speaking abilities can also influence the juror's acceptance of testimony. “If witnesses are articulate and confident in their assertions, jurors are likely to attach much more credibility to their testimony than if the witness speaks in a ‘powerless style.’” *Id.* at 147–48; see also Strier, *supra* note 68, at 114 (“Attention is too often focused on the personal characteristics of expert witnesses instead of the quality of their evidence.”).

76. See Jensen, *supra* note 54, at 187; Lunney, *supra* note 38, at 110. Jensen explains, “*Trial*, the magazine of the Association of Trial Lawyers, contains numerous experts’ advertisements proclaiming that they can bring in the highest judgments possible.” Jensen, *supra* note 54, at 187; see also Easton, *supra* note 65, at 465. Lamenting the current rules of expert witness testimony, Easton states, “In fact, it is difficult to imagine a system that would lead to more biased testimony.” Easton, *supra* note 65, at 471; see also Justin P. Murphy, Note, *Expert Witnesses at Trial: Where Are the Ethics?*, 14 GEO. J. LEGAL ETHICS 217 (2000). Murphy concludes, “Today, criticism of expert witnesses is widespread throughout the legal community. One can find and hire an expert to testify on virtually any topic, and even simple lawsuits often involve the testimony of an expert witness.” Murphy, *supra*, at 217–18 (citations omitted); see also L. Timothy Perrin, *Expert Witness Testimony: Back to the Future*, 29 U. RICH. L. REV. 1389, 1393 (1995). Perrin writes with a tone of disgust and outrage toward the expert witness system. He states, “The combination of zealous advocates, paid experts, liberal rules of admission, and untrained jurors raises the question of whether the adversary system produces a reliable and accurate evaluation of expert witness testimony, and whether it is capable of doing so.” Perrin, *supra*, at 1393; see also AM. MED. ASSOC., HOUSE OF DELEGATES PROCEEDINGS, 147TH ANNUAL MEETING 75–78 (June 14–18, 1998), [hereinafter AMA REPORT]. After stating that “medical witness[es] must not become an advocate or a partisan in the legal proceeding,” the AMA Board states that “[e]conomic incentives can color the nature of the physician expert’s testimony.” AMA REPORT, *supra*, at 77, 78; see also PETER W. HUBER, GALILEO’S REVENGE: JUNK SCIENCE IN THE COURTROOM 17 (1991). Objective truth might be the goal of the judicial process, but it is not the goal of hiring an expert witness. Huber quotes a former president of the American Bar Association who proclaimed, “I would go into a lawsuit with an objective uncommitted independent expert . . . about as willingly as I would occupy a foxhole with a couple of noncombatant soldiers.” HUBER, *supra*, at 18; see also Dick Thornburgh, *Junk Science—The Lawyer’s Ethical Responsibilities*, 25 FORDHAM URB. L.J. 449, 449 (1998). Former Attorney General Thornburgh gives a straightforward condemnation:

A look at the classified section of any legal publication will produce samples of a whole industry of “experts” advertising their abilities to

“junk science.”⁷⁷

Should the judge be responsible for evaluating expert testimony prior to trial? The current Federal Rules of Evidence require the judge to determine whether an expert may testify.⁷⁸ Some commentators

provide a wide range of expert testimony. Many of them get right to the point, highlighting jury awards or settlement amounts gained as a result of their testimony. . . . Their business is *litigation*, not science. Their motivation raises serious questions about the use of expert testimony generally. Are these experts really seeking to assist the trier of fact, or are they hired guns aiming at a pre-determined result?

Thornburgh, *supra*, at 452; see also Daniel W. Shuman et al., *An Empirical Examination of the Use of Expert Witnesses in the Courts—Part II: A Three City Study*, 34 JURIMETRICS J. 193, 205 (1994). Most experts are paid quite handsomely for their testimony. Shuman found that the average fee charged by expert witnesses was \$185 per hour, with a range from \$50 per hour to \$500 per hour. Shuman et al., *supra*, at 205; see also Perrin, *supra*, at 1414. Perrin concludes, “[N]o expert is immune from the bias that comes with compensation.” Perrin, *supra*, at 1415. Similarly, Jeffrey L. Harrison argues that the preference for effective rather than honest expert witnesses leads to social costs, including “wrong” decisions and excessive costs in courts’ attempts to expose expert biases. See generally Jeffrey L. Harrison, *Reconceptualizing the Expert Witness: Social Costs, Current Controls and Proposed Responses*, 18 YALE J. ON REG. 253 (2001). These costs, which are not internalized by experts themselves, must be absorbed in other parts of the judicial system. *Id.* at 259. Harrison argues that several controlling devices on experts, such as excluding expert testimony, “judicial shunning,” and legal action, are unlikely to reduce levels of dishonesty and bias testimony because these devices are not internalized by the experts themselves. *Id.* at 314. But see Miles J. Vigilante, Note, *Screening Expert Testimony After Kumho Tire Co. v. Carmichael*, 8 J.L. & POL’Y 543, 546 (2000) (acknowledging problems associated with the use of expert testimony, but nevertheless, arguing that the current rules governing the admissibility of expert testimony make for an adequate system); Scott E. Sundby, *The Jury as Critic: An Empirical Look at How Capital Juries Perceive Expert and Lay Testimony*, 83 VA. L. REV. 1109 (1997). In his discussion of juries’ perceptions of witnesses in capital cases Sundby concludes, “[E]xperts are only one of many tools necessary to build an effective case in mitigation.” *Id.* at 1188. Sundby also found juries reported negatively evaluating testimony of experts who they thought were biased. *Id.* at 1128. He found that defense experts in capital murder cases were more likely to be viewed unfavorably than their counterparts on the prosecution. *Id.* at 1123. Most importantly, his study of jurors found that experts were criticized for being hired guns and jurors were “skeptical of experts and their ability to explain human behavior.” *Id.* at 1125.

77. See, e.g., Browne et al., *supra* note 10, at 7; Thornburgh, *supra* note 76, at 467. Thornburgh suggests that “if . . . attorney[s] were to be held accountable for introducing evidence that later turns out to be junk science, attorneys would be less likely to risk the introduction of junk science.” Thornburgh, *supra* note 76, at 467. But see David S. Caudill, *Advocacy, Witnesses, and the Limits of Scientific Knowledge: Is There an Ethical Duty to Evaluate Your Expert’s Testimony?*, 39 IDAHO L. REV. 341, 354–55 (2003) (arguing that because the lawyer has a duty to act as a zealous advocate for her client, she should not be responsible for preventing junk science from coming into court). See generally HUBER, *supra* note 76 (decrying the presence of “junk science” in modern legal proceedings).

78. FED. R. EVID. 104. Rule 104(a) provides:

argue that the judge, as the “gatekeeper,” is more capable of evaluating the expertise than is the jury.⁷⁹ Depending on the amount of trust they put in their own evaluative capabilities regarding complex evidence and testimony, judges may be more or less flexible in admitting evidence.⁸⁰ While judges are permitted to request help in evaluating an expert witness’s testimony,⁸¹ some judges may simply rely on the adversarial nature of court proceedings to elicit reliable testimony or rely on the

Preliminary questions concerning the qualification of a person to be a witness, the existence of a privilege, or the admissibility of evidence shall be determined by the court, subject to the provisions of subdivision (b). In making its determination it is not bound by the rules of evidence except those with respect to privileges.

This requirement means that the judge must consider whether the expert is proposing to testify to scientific knowledge and whether that knowledge will assist the trier of fact. *See* FED. R. EVID. 702.

79. In *Daubert v. Merrell Dow Pharmaceuticals, Inc.*, the Court takes the position that judges are more capable than juries to evaluate expertise. 509 U.S. 579, 593 (1993). *But see id.*, at 599 (Rehnquist, C.J., concurring in part and dissenting in part) (discussing doubts that judges can understand areas of expertise in which they lack formal training); *see also* James J. Elacqua & L. Gene Spears, *Technical Trials Call for Unique Strategies*, NAT’L. L.J., June 6, 1994, at C2 (“A jury, when properly educated through trial witnesses, argument and instruction, is every bit as qualified as a judge to decide a technical case.”).

80. *See, e.g., Confronting the New Challenges*, *supra* note 37, at 1513. Some commentators suggest that when a judge feels uncertain about his or her ability to evaluate evidence, she may be more lenient in allowing the expert’s testimony. *See, e.g., Zuchowicz v. United States*, 870 F. Supp. 15, 19 (D. Conn. 1994).

Jurors, however, may not be aware of this self-doubt, and may tend to regard testimony admitted by the judge as endorsed thereby. *See* Easton, *supra* note 65, at 480.

Just in case the recitation of an expert’s credentials might not be enough to elevate her to a special status above fact witnesses, in many courtrooms this recitation is followed, sometimes even in the absence of an objection to the expert’s opinion testimony, by a judicial declaration that the witness is indeed an “expert” who is thereby specially qualified to “assist the trier of fact to understand the evidence or to determine a fact in issue.”

Id.

81. *See generally* Note, *Improving Judicial Gatekeeping: Technical Advisors and Scientific Evidence*, 110 HARV. L. REV. 941 (1997) (examining and arguing for the use of technical advisors by the courts to assist judges in weighing expert testimony). *See also Confronting the New Challenges*, *supra* note 37, at 1517. Institutions are creating texts to aid the judge in this task. For an argument based on optimism about judges’ abilities, *see* SHEILA JASANOFF, *SCIENCE AT THE BAR: LAW, SCIENCE, AND TECHNOLOGY IN AMERICA* 42, 68 (1995). Jasanoff suggests that judges should increase their scientific literacy to decrease the distorting power of expert witnesses. *Id.* at 68. For example, given the frequency of toxic tort cases, it may be extremely beneficial for judges to gain understanding of common medical procedures and studies including cancer bioassays, risk analysis, and epidemiological studies. *Id.*

scientific community to help determine the admissibility of evidence.⁸² In any case, the judge's general understanding of expertise, as well as the legal standards for regarding admissibility of expert testimony,⁸³ influence the judge's ability to fulfill the role of gatekeeper.⁸⁴

Experts themselves may have a significant role in this process of evaluating expert testimony. As we will see in our survey of the historical treatment of experts in American courts, the admissibility standards have been, at times, based heavily on a determination of "general acceptance"⁸⁵ within the specific discipline to which the expert belongs.⁸⁶ In other words, when the court allows testimony under the criterion of "acceptance," it is according deference to the determination of the experts themselves⁸⁷ about which testimony is appropriate.⁸⁸

82. See Kaushal B. Majmudar, *Daubert v. Merrell Dow: A Flexible Approach to the Admissibility of Novel Scientific Evidence*, 7 HARV. J.L. & TECH. 187, 195–96 (1993).

83. David L. Faigman, *Mapping the Labyrinth of Scientific Evidence*, 46 HASTINGS L.J. 555 (1995). When considering admissibility standards, Faigman argues that the judges must be "sophisticated consumers of science." *Id.* at 556. He believes this sophistication should come from understanding basic principles of statistics and research used by scientists. *Id.* at 558–59.

84. See, e.g., Allen & Miller, *supra* note 8, at 1144. Admitting or excluding evidence may determine a case's outcome. Wendy Fleishman & Russell Jackson, *Challenges to the Admissibility of Expert Testimony: What Works After Daubert?*, 723 P.L.I. Comm. 121 (1995); see also *Rosen v. CIBA-GEIGY Corp.*, 78 F.3d 316, 318 (7th Cir. 1996) (noting that excluding the expert's testimony regarding the role of a nicotine patch would have "doomed" the plaintiff's case).

85. The *Frye* standard offered the most comprehensive statement of this idea. See *infra* Parts II.B.–D. and accompanying text.

86. *Confronting the New Challenges*, *supra* note 37, at 1511 ("[T]he community of experts within a particular field is most qualified to assess the validity of an expert's theory or technique on that subject."); see also *United States v. Addison*, 498 F.2d 741, 743–44 (D.C. Cir. 1974).

87. A questionable idea, according to this quote by George Bernard Shaw: "All professions are conspiracies against the laity." See CHAFETZ, *supra* note 66, at 123 (citing GEORGE BERNAD SHAW, *THE DOCTOR'S DILEMMA* (1906)).

88. The reaction of experts to each other may be used in other more obvious or profitable ways. For one suggestion, see Samuel R. Gross, *Expert Evidence*, 1991 WIS. L. REV. 1113. One of Gross's many suggestions to improve the use of expert witnesses in the legal system is to have opposing counsels' experts present objections directly to each other, with time to "respond intelligently or even revise their positions." *Id.* at 1149. He believes that this adjustment would lead to clearer presentation and overall higher quality expert evidence. *Id.*; see also Perrin, *supra* note 76, at 1445–46; Strier, *supra* note 68, at 168–70 (stating that experts are better equipped to be fact finders than lay juries and discussing a number of ways in which individuals with expertise could play a nonpartisan role in the judicial system); Thornburgh, *supra* note 76, at 469 ("[T]he court should always reserve the right to refer disputes over alleged 'junk science' to an independent panel of experts, not to decide the question in controversy, but to assess the quality of the expertise as required under the 'gatekeeping' regimen of *Daubert*.").

A major focus of this Article will be questioning the epistemological assumptions about scientific knowledge that this role of the expert entails. Seeing how the various players in the courtroom drama interact with this expertise leads us to realize the importance of understanding the nature of expert testimony as we seek to optimize its use in legal controversies. As we will see in examining the evolution of the American legal system's guidelines for such testimony, this question of roles has been central to the various standards held at one time or another.

II. THE LAW REGARDING THE USE OF EXPERTS

A. *The Historical Treatment of Experts*

Historically, a witness was allowed to testify only to what he had personally seen or heard⁸⁹ and not to any inferences or opinions the witness made from his or her observations.⁹⁰ However, the expert was not constrained by this restriction.⁹¹ Although the expert was not permitted to testify to the ultimate issue in the case,⁹² expert testimony was generally admissible as long as it pertained to knowledge “not within the common knowledge of the layman.”⁹³ For the testimony he

89. A “witness” is defined, in part, as “one who, being present, personally sees or perceives a thing.” BLACK’S LAW DICTIONARY 1603 (6th ed. 1990).

90. Although the categories of “facts” and “opinions” do not have a bright yellow line painted between them to clearly discern their separate identities, this distinction in legal thought continues to this day and plays a large part in the debate over legal testimony. See, e.g., Allen & Miller, *supra* note 8, at 1132. By restricting the layperson from testifying to opinions and inferences, the jury could gain closer access to the pure facts. Likewise, experts, providing the court finds the “right” experts, are expected to dispense from the storehouse of scientifically established truths. The difficulty comes, as we will see, when competing scientific truth claims are brought to the courtroom. Jacobs, *supra* note 40, at 1089 n.27.

91. In the 1782 case of *Folkes v. Chadd*, Lord Mansfield equated an expert’s knowledge in his field with personal knowledge, reconciling this deviation from the personal knowledge requirement. (1782) 99 Eng. Rep. 589, 589–90 (K.B.).

Under the present Federal Rules of Evidence, a lay witness can testify to opinion or inferences only if the opinions or inferences are rationally based on the perception of the witness and helpful to create a clear understanding of the witness’s testimony or the determination of a fact in issue. FED. R. EVID. 701. An expert, however, is not subject to this restraint on opinion with some exceptions. Under the Rules, an expert may testify to opinions so long as it will assist the trier of fact or determine a fact in issue. FED. R. EVID. 702.

92. Lee Waldman Miller, Comment, *Cross-Examination of Expert Witnesses: Dispelling the Aura of Reliability*, 42 U. MIAMI L. REV. 1073, 1076 (1988).

93. *Id.* at 1074 (citing *Bridger v. Union Ry.*, 355 F.2d 382, 387 (6th Cir. 1966)). The Federal Rules of Evidence eliminated the requirement that limited expert testimony to

did offer, the expert had to state his qualifications, describe the facts underlying his opinions, and explain the basis for these opinions.⁹⁴ The fact finder could then evaluate whether the testimony was consistent with the data used by the expert.⁹⁵

B. The Frye Test

Under the common law, experts had to explain the facts underlying their opinions. One particularly common, yet confusing, type of evidence that experts have been asked to explain is scientific evidence. The courts have treated novel scientific evidence differently from other types of evidence since the 1923 Court of Appeals for the District of Columbia decision in *Frye v. United States*.⁹⁶ This case posed a confusing standard dictating that the science forming the basis of the expert testimony is to be “generally accepted” in the scientific community,⁹⁷ placing the burden of demonstrating this standard upon the expert.⁹⁸ In spite of the uncertainty as to how the courts should apply this standard,⁹⁹ the precedent remained in place for the following

matters outside a layperson’s understanding. See FED. R. EVID. 702 (requiring only that the testimony aid the trier of fact).

94. These facts had to be facts of which the expert had personal knowledge, eliminating the dangers of hearsay testimony, and the facts, data, and opinions presented by the expert had to be submitted as hypothetical questions. Miller, *supra* note 92, at 1074.

95. *Id.*

96. 293 F. 1013 (D.C. Cir. 1923). In *Frye*, the defendant attempted to offer expert witness testimony regarding the results of a systolic blood pressure deception test, a test similar to a polygraph machine. *Id.* at 1013. The expert claimed that the results of the test verified the defendant’s innocence. *Id.* at 1014.

97. Browne et al., *supra* note 10, at 10–11. The authors explain that this short, citation-free opinion was riddled with ambiguity. *Id.*; see also *Frye*, 293 F. at 1013–14. The court determined that the deception test had not gained sufficient recognition with physiological and psychological experts. *Frye*, 293 F. at 1014. According to Justice Van Orsdel,

Just when a scientific principle or discovery crosses the line between experimental and demonstrable stages is difficult to define. Somewhere in this twilight zone the evidential force of the principle must be recognized, and while courts will go a long way in admitting expert testimony deduced from a well-recognized scientific principle or discovery, the thing from which the deduction is made must be sufficiently established to have gained general acceptance in the particular field in which it belongs.

Id.

98. Miller, *supra* note 92, at 1093; Browne et al., *supra* note 10, at 11.

99. Miller, *supra* note 92, at 1093–94. It is clear, however, that the acceptance of one, let alone ten, qualified experts is not enough to make a scientific procedure admissible testimony under this standard. See *Confronting the New Challenges*, *supra* note 37, at 1486. Moreover,

fifty years.¹⁰⁰

C. Application of The Frye Test

The scope of the *Frye* decision is a matter of debate. The standard for admissibility was created in the criminal context in relation to scientific testimony, and most courts initially limited the decision to the criminal law setting.¹⁰¹ However, in 1984, *Frye* was first applied to the civil context.¹⁰² The aspect of testimony under evaluation by the court is both the scientific technique applied by the expert and the principles on which the technique rests.¹⁰³

if there is a three-way tie among experts, under the *Frye* standard, there is no general acceptance. *Id.* Professor McCormick has suggested a standard in which any relevant evidence supported by a qualified expert would be admissible, a standard perhaps even more permissive than the current standard. See Osborne, *supra* note 59, at 512 (citing CHARLES T. MCCORMICK, HANDBOOK OF THE LAW OF EVIDENCE 363-64 (1954)).

100. See Clayton C. Skaggs, *Evidence: Say Good-Bye to the Frye "General Acceptance" Test*, 33 WASHBURN L.J. 450, 450, 458 n.59 (1994); Miller, *supra* note 92, at 1075; Jon P. Thames, *It's Not Bad Law—It's Bad Science: Problems with Expert Testimony in Trial Proceedings*, 18 AM. J. TRIAL ADVOC. 545, 549 (1995).

The *Frye* standard has been applied to various types of scientific evidence. See *United States v. Tranowski*, 702 F.2d 668, 669 (7th Cir. 1983) (astronomer's testimony dating photographs by measuring lengths of shadows); *Hughes v. Mathews*, 576 F.2d 1250, 1258 (7th Cir. 1978) (psychiatric testimony); *United States v. Kilgus*, 571 F.2d 508, 510 (9th Cir. 1978) (forward looking infrared imaging systems); *Lindsey v. United States*, 237 F.2d 893, 894 (9th Cir. 1956) (sodium pentothal); *United States v. Addison*, 498 F.2d 741, 743 (D.C. Cir. 1974) (voice prints); *People v. Palmer*, 145 Cal. Rptr. 466, 472 (Ct. App. 1978) (scanning electron microscope); *People v. Slone*, 143 Cal. Rptr. 61, 68 (Ct. App. 1978) (bite-mark comparisons); *People v. Alston*, 362 N.Y.S.2d 356, 362 (Sup. Ct. 1974) (blood testing); *State v. Smith*, 362 N.E.2d 1239, 1246 (Ohio Ct. App. 1976) (gunshot residue tests).

Some of the areas to which the opinion has been applied are particularly remote. In one context, courts have prohibited experts from testifying that an accused child molester deviates from the profile of a child abuser because that testimony is infringing on the jurors' role of assessing the credibility of the abuser. See, e.g., *People v. Berrios*, 568 N.Y.S.2d 512, 513 (Sup. Ct. 1991) (refusing to admit profile testimony of a child abuser).

The *Frye* standard was applied by courts even several years after the enactment of the Rules. See, e.g., *United States v. Shorter*, 809 F.2d 54, 60 (D.C. Cir. 1987) (application of *Frye* standard subsequent to Rules' enactment); *United States v. Brown*, 557 F.2d 541, 542 (6th Cir. 1977) (applying *Frye* subsequent to enactment of the Rules).

101. See *Confronting the New Challenges*, *supra* note 37, at 1529 (citing Paul C. Giannelli, "Junk Science": *The Criminal Cases*, 84 J. CRIM. L. & CRIMINOLOGY 105, 111 (1993)); see also Huber, *supra* note 36, at 733.

102. See Paul C. Giannelli, "Junk Science": *The Criminal Cases*, 84 J. CRIM. L. & CRIMINOLOGY 105, 111 (citing *Barrel of Fun, Inc. v. State Farm Fire & Cas. Co.*, 739 F.2d 1028 (5th Cir. 1984)).

103. See Melissa M. Horne, Note, *Novel Scientific Evidence: Does Frye Require That General Acceptance Within the Scientific Community be Established by Disinterested Scientists?*, 65 U. DET. L. REV. 147, 154-55 (1987) (citing Paul C. Giannelli, *The Admissibility of Novel Scientific Evidence: Frye v. United States, a Half Century Later*, 80 COLUM. L. REV.

Thus, the *Frye* test can be seen to possess some advantages. The burden to evaluate complicated scientific evidence is removed from judge and jury,¹⁰⁴ consistency within the legal system may be attained, and questionable testimony can be more carefully excluded.¹⁰⁵ Defenders of the standard recognize that requiring generally accepted scientific evidence excludes testimony that may be valid.¹⁰⁶ However, they argue that the cost of excluding worthy testimony from the fringes is outweighed by the benefit of protecting the court from scientific charlatanism.¹⁰⁷

However, the confusion precipitated by the *Frye* standard was remarkable.¹⁰⁸ Depending on the court's stipulated meaning of "general

1197, 1211 (1980)).

104. Browne et al., *supra* note 10, at 14; John William Strong, *Language and Logic in Expert Testimony: Limiting Expert Testimony by Restrictions of Function, Reliability, and Form*, 71 OR. L. REV. 349, 366 (1992); Tamarelli, *supra* note 71, at 1179.

105. Browne et al., *supra* note 10, at 14. The authors point out that the jurors will be spared having to wade through "scientific banter" and "unpublished hunches," both of which can hinder the fact finder's quest for truth. *Id.* (citations omitted).

106. Tamarelli, *supra* note 71, at 1178–79; *see also* *Daubert v. Merrell Dow Pharms., Inc.*, 509 U.S. 573, 597 (1993) (acknowledging that some authentic insights will still be excluded under the new standard).

107. The *Frye* standard could let in so-called charlatans as well. *See Confronting the New Challenges*, *supra* note 37, at 1511 (citing HUBER, *supra* note 76, at 14); Osborne, *supra* note 59, at 509–11 (recognizing that unreliable evidence could satisfy the *Frye* standard, and the scientific community will not always extensively test a novel scientific technique). For example, the paraffin test, generally accepted in the scientific community as a way to detect nitrate residues from gunpowder on a person's hands, was found to create a high number of false positives (from tobacco or nail polish) after having been used to convict defendants for a quarter-of-a-century. *See* Jacobs, *supra* note 40, at 1088 (citing Edward J. Imwinkelried, *Judge Versus Jury: Who Should Decide Questions of Preliminary Facts Conditioning the Admissibility of Scientific Evidence?*, 25 WM. & MARY L. REV. 577, 580 (1984)). The problem of jurors splitting the difference between conflicting experts is related to a concern that jurors cannot distinguish between "charlatans and Nobel prize winners." *Id.* at 1084 n.8; *see, e.g.*, Tamarelli, *supra* note 71, at 1176; Osborne, *supra* note 59, at 508–09. *But see* *United States v. Baller*, 519 F.2d 463, 466 (4th Cir. 1975) (preferring the approach of permitting the jury to hear testimony and assess the credibility through cross-examination and refutation).

108. *See* *Campbell v. People*, 814 P.2d 1, 8 (Colo. 1991) (refusing to apply *Frye* to expert psychological testimony on eyewitness reliability because eyewitness testimony was not a scientific device); Hutchinson & Ashby, *supra* note 36, at 1907 n.157 ("Few decisions have dominated an area of the law or caused as much confusion as *Frye v. United States*." (quoting Bert Black, *A Unified Theory of Scientific Evidence*, 56 FORDHAM L. REV. 595, 629 (1988))). *But see* Frank R. Emmerich, Jr., Note, *The Supreme Court Strengthens the Discretionary Powers of the District Courts in Admitting Expert Scientific Testimony: Daubert v. Merrell Dow Pharmaceuticals, Inc.*, 3 WIDENER J. PUB. L. 1051, 1053–54 (1994) ("[T]he application of this new rule [the *Daubert* standard] will only lead to greater confusion and frustration than that associated with *Frye*.").

acceptance,”¹⁰⁹ evidence may be accepted in some courts while rejected in others.¹¹⁰ Furthermore, giving such a substantial role to the expert, some argue, deprives the parties of their right to trial by jury¹¹¹ and keeps them from presenting evidence that supports their case.¹¹² As a response to this confusion, the Federal Rules of Evidence were created.

D. The Federal Rules of Evidence

In 1975, Congress adopted the Federal Rules of Evidence.¹¹³ The Rules specifically addressed the admissibility of expert testimony, eliminating some of the barriers to the admission of expert testimony.¹¹⁴ Rule 401 defines “relevant evidence” as any evidence that makes the existence of a material fact more or less probable,¹¹⁵ while Rule 402 permits admissibility of all relevant evidence.¹¹⁶ Rule 104(a) requires the judge, rather than the jury, to determine whether an expert can testify.¹¹⁷ Rule 702 provides the specific guidelines for the judge’s decision, stating that a witness qualified as an expert may testify to scientific, technical, or specialized knowledge, in the form of opinion or

109. Tamarelli identified various judicial definitions of “general acceptance,” such as “widespread, prevalent, extensive though not universal.” See Tamarelli, *supra* note 71, at 1186–87.

110. See generally Eileen A. Scallen, *Classical Rhetoric, Practical Reasoning, and the Law of Evidence*, 44 AM. U. L. REV. 1717 (1995).

111. U.S. CONST. amend. VII.

112. See, e.g., Paul C. Giannelli, *The Admissibility of Novel Scientific Evidence: Frye v. United States, a Half Century Later*, 80 COLUM. L. REV. 1197, 1230–31 (1980). Stringently excluding expert testimony may infringe on the defendant’s right to present exculpatory evidence; however, too much flexibility in admitting evidence may prejudice the jury. *Id.*

113. Enacting the rules took over thirteen years. S. Rep. No. 93-1277 (1974), *reprinted in* 1974 U.S.C.C.A.N. 7051, 7051–52.

114. See, e.g., *Holbrook v. Lykes Bros. S.S. Co.*, 80 F.3d 777, 781 (3rd Cir. 1996) (noting Rule 702’s liberal policy of admission); see Lunney, *supra* note 38, at 105.

115. FED. R. EVID. 401; see *Daubert v. Merrell Dow Pharms., Inc.*, 509 U.S. 579, 587 (1993) (“The Rule’s [Rule 401] basic standard of relevance thus is a liberal one.”).

116. FED. R. EVID. 402 (“All relevant evidence is admissible, except as otherwise provided by the Constitution of the United States, by Act of Congress, by these rules, or by other rules prescribed by the Supreme Court pursuant to statutory authority. Evidence which is not relevant is not admissible.”).

117. FED. R. EVID. 104. Rule 104(a) provides:

Preliminary questions concerning the qualification of a person to be a witness, the existence of a privilege, or the admissibility of evidence shall be determined by the court, subject to the provisions of subdivision (b). In making its determination it is not bound by the rules of evidence except those with respect to privileges.

otherwise,¹¹⁸ if it “will assist” the trier of fact.¹¹⁹ If an expert bases her opinion on facts and data that are reasonably relied upon by other experts in their fields, Rule 703 permits the admission of those facts or data.¹²⁰ Finally, according to Rule 706, the judge has the discretion to appoint an expert for assistance.¹²¹ Rule 706 has particular salience for

118. The Rules eliminated the requirement that counsel use hypotheticals in certain circumstances, *see* FED. R. EVID. 705, and abolished the restriction against testifying to the ultimate issue, *see Confronting the New Challenges, supra* note 37, at 1486–87. At common law, an expert could testify that something could have caused a particular result but not that the expert believed it did cause that outcome in the case at hand. Miller, *supra* note 92, at 1076.

Despite the allowance of testimony on the ultimate issue in the case, most judges still keep the expert from taking that final step, preferring to leave the jury to draw its own conclusions on the implications of the expert’s testimony. *See* Weinstein, *supra* note 45, at 717.

Congress retained one exception to the lifting of the ultimate issue ban in Rule 704(b), which forbids an expert from testifying as to the mental state or condition of a criminal defendant, specifically as to whether the defendant had the mental state or condition constituting an element of the crime charged. FED. R. EVID. 702.

119. FED. R. EVID. 702 (“If scientific, technical, or other specialized knowledge will assist the trier-of-fact to understand the evidence or to determine a fact in issue, a witness qualified as an expert by knowledge, skill, experience, training, or education, may testify thereto in the form of an opinion or otherwise.”).

Professor Frederic I. Lederer, holding that reliability of testimony is the universal underlying concern, proposed the insertion of the word “reliable” before “scientific” in Rule 702. *See* Frederic I. Lederer, *Proposals for a Model Rule on Admissibility of Scientific Evidence: Resolving the Frye Dilemma—A Reliability Approach*, 115 F.R.D. 79, 84 (1987).

120. *See* FED. R. EVID. 703. Some scholars argue this provision incorporates the *Frye* standard into the Rules by appealing to the mainstream expert community view. *See, e.g.,* Laurel Beeler & William R. Wiebe, Comment, *DNA Identification Tests and the Courts*, 63 WASH. L. REV. 903, 937 n.176 (1988); Gordon J. Beggs, *Novel Expert Evidence in Federal Civil Rights Litigation*, 45 AM. U. L. REV. 1, 34 n.246 (1995).

121. Rule 706(a) provides:

Appointment. The court may on its own motion or on the motion of any party enter an order to show cause why expert witnesses should not be appointed, and may request the parties to submit nominations. The court may appoint any expert witnesses agreed upon by the parties, and may appoint witnesses of its own selection. An expert witness shall not be appointed by the court unless the witness consents to act. A witness so appointed shall be informed of the witness’ duties by the court in writing, a copy of which shall be filed with the clerk, or at a conference in which the parties shall have an opportunity to participate. A witness so appointed shall advise the parties of the witness’ findings, if any; the witness’ deposition may be taken by any party; and the witness may be called to testify by the court or any party. The witness shall be subject to cross-examination by each party, including a party calling the witness.

See Margaret G. Farrell, *Coping with Scientific Evidence: The Use of Special Masters*, 43 EMORY L.J. 927 (1994), for a detailed discussion of the use of court-appointed witnesses. For

this Article because it provides the judge with an opportunity to better use social science expertise.

The Rules are considered more liberal than the *Frye* test;¹²² however, *Frye* was not mentioned anywhere within the Rules. Thus, scholars debated whether the Rules took precedence over the *Frye* test. Those who believed that *Frye* was overruled held that the drafters of the Rules would have explicitly mentioned *Frye* had it remained applicable to expert testimony.¹²³ Those who held that *Frye* continued to be valid argued that the drafters would have explicitly mentioned *Frye* as being overruled¹²⁴ and that the Rules were not a comprehensive statement of

further discussion of court-appointed experts, see, for example, Strier, *supra* note 68, at 115. Strier states, “[E]xpert testimony is today almost always confined to those experts hired by the parties—often to the detriment of the factfinder.” *Id.* Strier goes on to say, “No question exists as to the judge’s authority to call expert witnesses.” *Id.* at 177. He speculates that court appointed expert witnesses are not used frequently because judges fear they would be too influential. *Id.* at 177–78. Strier tries to allay fears of excessive influence by suggesting that judges could tell jurors not to assume that court appointed experts had their testimony endorsed by the court. *Id.* at 178. However, the tendency to treat court appointed experts’ testimony with superior reverence seems likely. See, e.g., AMA REPORT, *supra* note 76. “If it is difficult for the judge to evaluate the evidence, the [c]ourts should be encouraged to exercise a power that they have, but seldom use: the ability to retain their own nonpartisan expert under Fed. R. of Evid. 706.” *Id.*; see also Deason, *supra* note 3, at 61. Deason recognizes that appointed experts can help juries and judges understand specialized issues. *Id.* at 81–94. However, she points out that eliminating payment to expert witnesses does not eliminate all biases that an individual carries. *Id.* at 112 (“An expert’s financial interests may be the most straightforward to identify and even regulate, but they may not be as influential as intellectual or other personal motivations.”); see also Vigilante, *supra* note 76, at 588–89 (claiming that court appointed expert witnesses are still biased and have too much power).

Germany similarly uses court-appointed experts in certain cases. See Sven Timmerbeil, *The Role of Expert Witnesses in German and U.S. Civil Litigation*, 9 ANN. SURV. INT’L & COMP. L. 163, 163 (2003). If both parties use experts and the experts disagree about the evidence, the German judge is required to hire a court expert who must testify. *Id.* at 178. Additionally, partisan experts’ testimony is not considered evidence in Germany. *Id.* at 178. Timmerbeil argues that this use of court experts may unduly influence judges and juries such that the court expert becomes the “de facto decision-maker.” *Id.* at 182. Furthermore, Timmerbeil argues that German judges may exert greater influence than American judges because German judges must choose the court expert. *Id.* at 185.

122. See *Confronting the New Challenges*, *supra* note 37, at 1487.

123. See *U.S. v. Downing*, 753 F.2d 1224, 1234 (3d Cir. 1985) (quoting Jack Weinstein & Margaret Berger, 4 WEINSTEIN’S FEDERAL EVIDENCE § 702[03], at 702–12 n.6); *U.S. v. Williams*, 583 F.2d 1194, 1198 (2d Cir. 1978). The Third Circuit stated that the *Frye* test was out. *Downing*, 753 F.2d at 1237. The Second Circuit, decrying the simplistic head-counting of *Frye*, held that courts should consider the potential rate of error of the expert’s scientific technique, the standard, or lack thereof, controlling the given technique, how the technique was used, and whether the technique was subject to abuse. *Williams*, 583 F.2d at 1198.

124. Congressional history, including the Advisory Committee Notes, floor debates, hearings, and committee reports regarding Rule 702, does not indicate whether Congress’s intent was to eliminate *Frye*. Tamarelli, *supra* note 71, at 1182.

common law guidelines for evidence.¹²⁵ A majority of circuit courts interpreted the two standards as coexistent,¹²⁶ until the 1993 Supreme Court ruled otherwise in *Daubert v. Merrell Dow Pharmaceuticals, Inc.*¹²⁷

E. Daubert: A New Standard of Admissibility

In *Daubert*, each party's case rested on expert testimony, and the conflict between the two scientific viewpoints hinged upon the court's admission of this testimony. Two children brought suit against Merrell Dow, alleging that its drug, Bendectin, caused their limb reduction birth defects.¹²⁸ Merrell Dow produced a "well-credentialed expert" who testified that based on his review of more than thirty published studies regarding Bendectin and human birth defects the causal link was not supported by any study.¹²⁹ In response, the plaintiffs' eight experts¹³⁰ testified that test tube and animal studies demonstrated a link between the drug and defects.¹³¹

Relying on circuit court decisions, the district court reasoned that the animal study evidence was inadmissible because it was not generally accepted in its field.¹³² Furthermore, the epidemiological analyses by the plaintiff's experts of previously published studies were inadmissible because the results of those analyses had not been published or subjected to peer review.¹³³

125. Horne, *supra* note 103, at 153; *see also* Tamarelli, *supra* note 71, at 1185 (stating that Rule 702 did not contain "a completely integrated standard of admissibility").

126. *See, e.g.*, Christophersen v. Allied-Signal Corp., 939 F.2d 1106, 1110 (5th Cir. 1991) (assuming the *Frye* standard survived the Rules); United States v. Smith, 869 F.2d 348, 350–51 (7th Cir. 1989).

127. 509 U.S. 579 (1993).

128. *Daubert v. Merrell Dow Pharms., Inc.*, 727 F. Supp. 570, 571 (S.D. Cal. 1989). Doctors prescribed Bendectin to pregnant women to reduce nausea.

129. *Daubert*, 509 U.S. at 582 ("[There were] more than 30 published studies involving at least 130,000 patients.").

130. These experts had the following credentials: Chief of the California Department of Health and Services Division, Senior Science Advisor to the Environmental Protection Agency and consultant to the Food and Drug Administration on causation issues, Professor of Epidemiology at the University of Texas, Associate Professor of Pediatrics, Pharmacology and Toxicology at the University of Texas Medical Branch, Professor of Pediatrics and Pharmacology at Wayne State University College of Medicine, Specialist in Pathology and Pharmacology, and director of the multi-discipline teaching labs and an Assistant Professor of pharmacology at the University of Arizona College of Medicine. *Id.* at 583 n.2.

131. The scientists believed that Bendectin ingestion while the fetus's limbs were forming could interfere with development. *Id.* at 583.

132. *Id.* at 583–84.

133. *Daubert*, 727 F. Supp. at 575. Various circuit courts refused to admit reanalysis

The Ninth Circuit affirmed the district court's decision by ruling that the test tube studies were not sufficient to provide a basis for causation.¹³⁴ The Supreme Court heard the case to determine whether the *Frye* test had survived the enactment of the Rules.¹³⁵

In *Daubert*, the plaintiffs argued that the Rules superseded the *Frye* rule; specifically, they claimed that Rule 402 admitted all relevant evidence.¹³⁶ In contrast, Merrell Dow, the defendant, suggested that Rule 702 created a general acceptance standard.¹³⁷ The Supreme Court rejected the traditional test for admissibility.¹³⁸ They decided that instead of relying on external groups to determine validity, judges themselves should make the decisions.¹³⁹ The Court ruled that evidence could be admitted under the Rules even if its scientific basis was not generally accepted;¹⁴⁰ thus, the Court settled the controversy that the Rules displaced the *Frye* test.¹⁴¹ However, the Court determined that judges needed to admit evidence only if it is both relevant and reliable.¹⁴² Some have suggested that this new standard, reliant upon the

studies of Bendectin risks because those studies had not been subjected to peer review. *Daubert v. Merrell Dow Pharms., Inc.*, 951 F.2d 1128, 1130–31 (9th Cir. 1991).

134. Three circuits had determined that animal and chemical studies were insufficient to demonstrate the link between Bendectin and birth defects. *Id.* at 1130 (citing *Brock v. Merrell Dow Pharms., Inc.*, 874 F.2d 307 (5th Cir. 1989), modified by 884 F.2d 166 (5th Cir. 1989); *Richardson v. Richardson-Merrell, Inc.*, 857 F.2d 823 (D.C. Cir. 1988); *Lynch v. Merrell-National Labs*, 830 F.2d 1190 (1st Cir. 1987)).

135. *Daubert*, 509 U.S. at 585. The Court's example of the conflict among the circuits is the decisions of *United States v. Shorter*, 809 F.2d 54, 59–60 (D.C. Cir. 1987) and *DeLuca v. Merrell Dow Pharmaceuticals, Inc.*, 911 F.2d 941 (3d Cir. 1990). *Daubert*, 509 U.S. at 586. In *Shorter*, the D.C. Circuit Court of Appeals used the *Frye* test to evaluate an expert's testimony regarding an asserted compulsive gambling disorder defense to income tax evasion. *Shorter*, 809 F.2d at 59–60. In contrast, the Third Circuit in *DeLuca*, following a Rules derived standard, considered the following: "(1) the soundness and reliability of the process or technique used in generating the evidence, (2) the possibility that admitting the evidence would overwhelm, confuse, or mislead the jury, and (3) the proffered connection between the scientific research or test result to be presented, and particular disputed factual issues in the case." *DeLuca*, 911 F.2d at 954–55.

136. *Daubert*, 509 U.S. at 587.

137. See Emmerich, *supra* note 108, at 1061 & n.47 (citing Brief for Respondent at 9, *Daubert* (no. 92-102)).

138. *Daubert*, 509 U.S. at 587.

139. *Id.* at 589.

140. *Id.* at 588; see also Faigman, *supra* note 83, at 563 ("Daubert's replacement of *Frye* anticipates that there will be many cases in which the research lacks general consensus in its field, but it is accurate enough to assist the trial process.").

141. *Daubert*, 509 U.S. at 588. The Court supported this decision by noting that nowhere in the Rules, or the drafting history, was *Frye* mentioned. *Id.*

142. *Id.* at 591–92. To assist trial judges in determining relevancy and reliability, the Supreme Court offered four considerations. *Id.* at 593–94. First, did the expert use the

judge's determination,¹⁴³ may exclude as much or more relevant evidence than did *Frye*: the judge cannot admit any testimony unless its scientific validity and connection to the scientific field can be established.¹⁴⁴ Moreover, the judge must be able to evaluate the scientific principles and methodology of the experts in question.¹⁴⁵

F. After Daubert

There have been a multitude of cases concerning, involving, struggling with, or contemplating the *Daubert* ruling. However, one theme has emerged: confusion. Since the *Daubert* decision in 1993, the Supreme Court has addressed *Daubert* and Federal Rule of Evidence 702 only four times.¹⁴⁶ The most important of these four decisions was

scientific method defined as forming and testing a hypothesis? *Id.* at 593. Second, has the expert's theory or technique been subjected to peer review and publication? *Id.* Third, does the scientific technique have a significant rate of error? *Id.* at 594. Finally, is the methodology generally accepted in the relevant community? *Id.* Another possible question to aid determination of admissibility might be the following: Did the expert conduct any research with this litigation in mind? *See Confronting the New Challenges*, *supra* note 37, at 1515 (citing *Daubert*, 509 U.S. at 1317). According to the Court in *Daubert*, other rules must bear on this inquiry as well where applicable. 509 U.S. at 595 (citing to Rule 706 allowing for court experts and Rule 403's balancing test). Additionally, Rule 703 allows for hearsay if the opinion is "of a type reasonably relied upon by experts . . . in forming opinions or inferences upon the subject." FED. R. EVID. 703.

143. The Supreme Court in *Daubert* first found that the *Frye* Court's exclusive reliance on the general acceptance standard was improper. 509 U.S. at 588. For a case admitting testimony not "generally accepted" in the community under *Daubert*, see *FDIC v. Suna Assocs. Inc.*, 80 F.3d 681 (2d Cir. 1996).

144. *See, e.g.*, Lunney, *supra* note 38, at 141; Emmerich, *supra* note 108, at 1106 (describing the *Daubert* standard as more conservative than *Frye*).

145. *Daubert*, 509 U.S. at 595; *see also Confronting the New Challenges*, *supra* note 37, at 1513. *But see Daubert*, 509 U.S. at 598–601 (Rehnquist, C.J., concurring in part and dissenting in part) (noting that judges are trained in interpretation of cases and statutory provisions, not interpretation of scientific knowledge, method, and validity). One scholar has noted the tension between the role of reliance upon evidence and the evaluation of that same evidence. *See Confronting the New Challenges*, *supra* note 37, at 1510. A possible improvement on the American role of judge as evaluator would be the British practice where complex matters are tried under "lay judges" with particular technical expertise. *See Warren E. Burger, Agenda for Change*, 54 JUDICATURE 232, 235 (1971).

146. *General Electric Co. v. Joiner*, 522 U.S. 136, 141–42 (1997). This case is generally cited for the Court's holding as to guidelines for appellate review of *Daubert* rulings. According to the Court, decisions concerning admitting expert testimony can only be reviewed under an "abuse of discretion" standard. *Id.*

Following the *Joiner* decision, the Supreme Court was compelled to indirectly address the *Daubert* ruling in *United States v. Scheffer*, 523 U.S. 303 (1998), regarding the admissibility of polygraph evidence. Here, the Court described some of the limits to *Daubert's* flexibility and, in very vague terms, attempted to define reliability. *Id.* at 312.

The third case to reference the *Daubert* decision is *Kumho Tire Co. v. Carmichael*, 526

*Kumho Tire Co.*¹⁴⁷ The Court concluded that the *Daubert* ruling

U.S. 137 (1999). The fourth time the Court referred to *Daubert* was in *Weisgram v. Marley Co.*, 528 U.S. 440 (2000). *Weisgram* informed the appellate courts as to what remedies they could use when reviewing *Daubert* appeals. *Id.* at 446. The Court held that appellate courts may remand a case, direct it for a new trial, direct the district court to decide if a new trial is warranted, or if the evidence is insufficient to a party's claim, the court may enter a judgment as a matter of law. *Id.* at 457. The Court opined, "We adhere to *Neely's* holding and rationale, and today hold that the authority of courts of appeals to direct the entry of judgment as a matter of law extends to cases in which, on excision of testimony erroneously admitted, there remains insufficient evidence to support the jury's verdict." *Id.* at 457.

Weisgram won a jury verdict. However, during the trial and in the post-trial motions, Marley repeatedly objected to certain evidence as inadmissible. *Id.* at 456. "[A]lthough *Weisgram* was on notice every step of the way that Marley was challenging his experts, he made no attempt to add or substitute other evidence." *Id.* The Court, citing *Lujan v. National Wildlife Federation*, 497 U.S. 871, 897 (1990), notes, "[A] litigant's failure to buttress its position because of confidence in the strength of that position is always indulged in at the litigant's own risk." *Weisgram*, 528 U.S. at 456.

The Eighth Circuit Court of Appeals agreed with Marley, ruling that the district court had erred in admitting *Weisgram's* expert testimony. *Id.* at 440–41. Subsequently, the court entered judgment for Marley as a matter of law. *Id.* *Weisgram* appealed, contending that entering a verdict as a matter of law for a loser in a jury trial was not within the scope of the appellate court's authority. *Id.* at 448–49. Specifically, the Court ruled that Federal Rule of Civil Procedure 501(a) can be applied to *Daubert* appeals. *Id.* at 445–46. The appeals court can direct a judgment as a matter of law for the loser of a jury trial if the court deems that evidence used by the other party, the winner, was erroneously admitted. The Court stated specifically,

But if, as in the instant case, the court of appeals concludes that further proceedings are unwarranted because the loser on appeal has had a full and fair opportunity to present the case, including arguments for a new trial, the appellate court may appropriately instruct the district court to enter judgment against the jury-verdict winner. Appellate authority to make this determination is no less when the evidence is rendered insufficient by the removal of erroneously admitted testimony than it is when the evidence, without any deletion, is insufficient.

Id. at 444.

However, a judgment as a matter of law can only be entered if the remaining evidence properly admitted does not support a decision for that party. *Id.* at 457.

Since *Daubert*, . . . parties relying on expert evidence have had notice of the exacting standards of reliability such evidence must meet. It is implausible to suggest, post *Daubert*, that parties will initially present less than their best expert evidence in the expectation of a second chance should their first try fail.

Id. at 455 (citations omitted). Circuit courts are granting summary judgments after the opposing party's expert testimony was expelled as inadmissible. Similarly, this case is cited in numerous circuit court decisions to quiet the objection that the court should not grant summary judgment but rather remand for a new trial. See, e.g., *Alfred v. Caterpillar, Inc.*, 262 F.3d 1083 (10th Cir. 2001).

147. 526 U.S. 137 (1999).

extended to non-scientific expert testimony.¹⁴⁸ Specifically, trial courts must judge the admissibility and reliability of all expert testimony in accordance with Federal Rule of Evidence 702.¹⁴⁹ Thus, *Kumho Tire Co.* effectively incorporated many of the soft sciences,¹⁵⁰ even those at one time considered junk science. Despite this guidance, circuit and district courts have been battling to understand the evidence presented in order to determine reliability to properly perform their “gate keeping” function.

The circuit courts have encountered problems interpreting and applying *Daubert*. A major obstacle appears to be a general lack of understanding concerning *Daubert* and its principles. Certain courts are taking the principles as rigid standards of admission that have effectively raised the bar for expert testimony, while others view *Daubert* as a flexible set of principles aimed at allowing the novel and respected minority expert opinions into the courtroom. On occasion the courts have varied significantly enough that they directly contradict one another.

Yet another problem that seems to be evident is the court’s ability to judge the merits of technical testimony. As gatekeepers for the court, judges are forced to evaluate the merits of testimony that is obviously beyond the purview of any court.¹⁵¹ Oftentimes the judges listen to the

148. *Id.* at 138.

149. *Id.*

150. *See, e.g.,* *Milanowicz v. Raymond Corp.*, 148 F. Supp. 2d 525 (D.N.J. 2001). The court, discussing the *Daubert* difficulties regarding *Kumho Tire*, states, “That guidance notwithstanding, in the wake of *Kumho*, courts have nevertheless struggled, with varying degrees of success, to apply the *Daubert* factors to seemingly recalcitrant technical subjects such as engineering.” *Id.* at 531; *see also* Kimberly M. Hrabosky, Note, *Kumho Tire v. Carmichael: Stretching Daubert Beyond Recognition*, 8 GEO. MASON L. REV. 203, 222 (1999) (arguing that “mechanical application of *Daubert* to nonscientific expert testimony defies logic and could lead to substantial injustice in the criminal context”); *cf.* Erica Beecher-Monas, *The Heuristics of Intellectual Due Process: A Primer for Triers of Science*, 75 N.Y.U. L. Rev. 1563, 1566–68 (2000) (arguing that judges gloss over genuine scientific inquiry due to perplexity over required inquiry).

151. *See* *Nelson v. Tenn. Gas Pipeline Co.*, 243 F.3d 244, 252 (6th Cir. 2001).

If anything, *Kumho* supports the magistrate judge’s consideration of factors not mentioned by the Supreme Court, including the fact that Kilburn’s study was conducted and the experts’ opinions were formed for purposes of litigation. This factor is consistent with our observation that close judicial analysis of expert testimony is necessary “because expert witnesses are not necessarily always unbiased scientists.” Here, the magistrate did not abuse his discretion by considering this factor as he did, or by concluding that “the fact that the study was performed in connection with litigation and funded by plaintiffs’ counsel does not militate in Dr.

arguments made by each side's witnesses about the validity of technical testimony and then are forced to apply vague guidelines enumerated by the Supreme Court in *Daubert* that may not be appropriate.

Courts also vary in their perceptions of their role under *Daubert*. The Ninth Circuit, for example, cited the district courts' "gatekeeper" role as "excluding 'junk science' that does not meet the standards of reliability required under Rule 702."¹⁵² In *Domingo v. T.K.*, the court held that if an expert does not test his own hypothesis, there are several other procedures the expert can employ to demonstrate validity and reliability.¹⁵³ "[I]f an expert did not conduct his or her own research, independent of the litigation, on the subject of the testimony, the district court must determine whether there exists any 'objective, verifiable evidence that the testimony is based on "scientifically valid principles.'""¹⁵⁴ This interpretation offers multiple methods other than testing to demonstrate reliability and validity, thereby allowing more expert testimony to be admitted.

The courts also have conflicting interpretations as to the extent to which an expert's qualifications can make up for gaps or assumptions in the reasoning. In *Campbell v. Metropolitan Property & Casualty Insurance Co.*,¹⁵⁵ the court did not require the expert to make scientific tests of his hypothesis but instead concluded that his assumptions were validated based on the credentials he possessed.¹⁵⁶ Based on his résumé,

Kilburn's favor."

Id. (citations omitted).

152. *Domingo v. T.K.*, 289 F.3d 600, 605 (9th Cir. 2002) (citation omitted).

153. *Id.* The court states,

On remand from the Supreme Court in *Daubert*, this court explained that, if an expert did not conduct his or her own research, independent of the litigation, on the subject of the testimony, the district court must determine whether there exists any "objective, verifiable evidence that the testimony is based on 'scientifically valid principles.'" Here, because Dr. Harrington had not conducted his own independent research on FES, the court correctly looked for objective and verifiable evidence of the validity of his theory.

Id. at 605 (quoting *Daubert v. Merrell Dow Pharms., Inc. (Daubert II)*, 43 F.3d 1311, 1317–18 (1995)).

154. *Id.* (quoting *Daubert II*, 43 F.3d at 1317–18).

155. 239 F.3d 179 (2nd Cir. 2001). In this case, the expert was asked to testify as to the time period in which lead poisoning began to affect the Campbell children. *Id.* at 179. Based on the impressive credentials of the expert witness the court granted him significant leeway. *Id.* at 186.

156. The court noted his credentials in responding to the claim his testimony was not

the court felt any inconsistencies in his method went to weight rather than admissibility, thus permitting him the benefit of the doubt.¹⁵⁷ Other courts have ruled that an extensive listing of credentials does not justify unsupported leaps of logic. The latter opinion can be seen in *Coffey v. Dowley Manufacturing, Inc.*,¹⁵⁸ where the court found that it must look beyond the expert's extensive pedigree.¹⁵⁹ Finding that his methods required taking his word as gospel, the court deemed his testimony inadmissible under *Daubert*.¹⁶⁰

In summary, depending on the court, circuit or district, expert testimony may be subject to radically different analyses.¹⁶¹ The role of

valid. *Id.* at 185–86.

He wrote several sections in the air-lead-criteria document published by the EPA, and he was chairperson of the Centers for Disease Control's advisory committee on the reports "published in '85 and '91 on preventing lead poisoning in young children, which are used nationally by departments of health as well as pediatricians for the treatment, diagnosis and management of childhood lead poisoning." Thus, the court noted that Dr. Rosen "seems to be a preeminent expert in the field relied on by all the relevant government agencies to establish the science for the policies that the government has adopted."

Id. at 186. (emphasis omitted).

157. *Id.* at 186. The court stated,

To the extent that Metropolitan asserts that there were gaps or inconsistencies in the reasoning leading to Dr. Rosen's opinion in this case, or that there were responses from which it could be argued that the onset of the children's injuries did not occur prior to the end of the first policy period, such arguments go to the weight of the evidence, not to its admissibility.

Id.

158. 187 F. Supp. 2d 958 (M.D. Tenn. 2002).

159. *Id.* at 975. "Yet, although Dr. Wilson's pedigree is impressive, the Rule 702 and *Daubert* analysis demands that [the] [c]ourt not simply take Dr. Wilson's bare assertions as gospel." *Id.* "[N]othing . . . requires a district court to admit opinion evidence which is connected to existing data only by the ipse dixit of the expert." (quoting *Gen. Elec. Co. v. Joiner*, 522 U.S. 136, 146 (1997)) (omission in original).

160. *Id.*

161. See Note, *Reliable Evaluation of Expert Testimony*, 116 HARV. L. REV. 2142, 2142–43 (2003). The author argues that because *Daubert* gave judges such little guidance, litigants will be tried inconsistently across circuits, and these inconsistencies may occur within the same trial court. *Id.* at 2147. To remove these inconsistencies, the author proposes that Congress create a taxonomic system for classifying various kinds of experts. *Id.* at 2143. With a classification system, judges could more easily and consistently evaluate the reliability of expert testimony. Here, the author suggests a two-tier system, composed of more experimental testimony and testimony that is based more on experience. *Id.*

the players in trial proceedings is still an open question: those advocating liberal admissibility standards tend to give a larger role to the jury,¹⁶² while those advocating increased restrictions¹⁶³ rely upon the judge to evaluate testimony before it is heard in trial.¹⁶⁴ In the next Part, we will examine the assumptions about the nature and origins of experts to see more clearly what it is that experts have to offer for the court's consumption.

III. EPISTEMOLOGICAL UNDERPINNINGS AND THE REAL STORY OF EXPERTISE

As we have seen in examining the various admissibility standards offered by American common law over the years, different views about the nature of expertise are implied by these standards.¹⁶⁵ Evaluating the

162. The jury may or may not be aware of the disputed nature of this evidence. See Perrin, *supra* note 76, at 1399. Perrin says,

Information that once would have been brought out by a lawyer on cross examination and argued during closing argument is now brought out through an expert witness who has access to all the testimonial advantages the Rules provide to experts. The expert can even testify to opinions that embrace the ultimate issue, thus enhancing the advocate role of the expert.

Id.; see also Scallen & Wiethoff, *supra* note 44, at 1151 (“[E]xpert witnesses do not have to reveal the bases for their opinions.”).

163. One state's regulations (Washington) provide examples of such restrictions. See Thornburgh, *supra* note 76, at 462.

164. See Scallen & Wiethoff, *supra* note 44, at 1145–46. The authors state,

[T]he problem of expert testimony cannot be resolved adequately simply by turning judges into “amateur scientists.” Instead, the problem . . . must be addressed by confronting the problem at bottom: Science and law speak different languages. A judge, even one trained in the scientific method, cannot alone translate the testimony of expert witnesses into meaningful legal discourse.

Id. They explain that now judges, rather than juries, are given primary responsibility for evaluating an expert's ethos, “thus obscuring the fact that expert testimony is, at bottom, ‘opinion,’ traditionally evaluated by the trier of fact, often a jury.” *Id.* at 1144. But see Abramson, *supra* note 51. Abramson argues, “[T]he scientific method and the law of evidence converge to place scientific gatekeeping squarely within the realm of judicial competence.” *Id.* at 726.

165. See, e.g., Robert Kargon, *Expert Testimony in Historical Perspective*, 10 LAW & HUM. BEHAV. 15, 15 (1986) (offering a general argument that our reverence for “experts” has changed significantly throughout history); Haskell, *supra* note 5, at xii (“[T]he noun form of the word [expert]—with all it implies about the distinctiveness of the social role, its visibility across a wide spectrum of activities and occupations, and the prospects of earning an income from it—did not come into use until the middle decades of the nineteenth century.”). For

standards that have been presented, as well as the often-conflicting judicial application of these standards, leads to important questions about the epistemological assumptions inherent to these different views.¹⁶⁶ How do experts arrive at the knowledge they present for our edification? Are scientific “facts”¹⁶⁷ as we receive them the products of an impersonal and objective scientific force,¹⁶⁸ or the results of human

further discussion of the development of our current reliance upon expertise in the twenty-first century, see BRINT, *supra* note 4, at 40 (“[T]he prominence of expert professionalism grew naturally out of the sense of progress and power long associated with new technology and new intellectual tools of control.”). See also Haskell, *supra* note 5, at xii (“[A]scending levels of population density and per capita income made it possible for substantial numbers of people to make a living by selling advice and specialized services, rather than producing food or other tangible goods.”). But see Larson, *supra* note 5, at 37 (“[P]owerful as they were, the legitimizing factors that we have discussed so far—faith in science, confirmed by technological progress; the submission of all to the impersonal laws of the market, rigged though it was—were not enough to confer upon experts the full scope and character of the power they appear to exercise today.”). Instead, Larson argues that the rise of expert power should largely be contributed to the “emergence of a formally free system of mass education.” *Id.* To Larson, it was the “pedagogies based on an ‘ideal of simple calculability’” that reinforced the notion that “experts, who study longer, master uncertainty better because they know *more facts*.” *Id.* at 55. “Expertise . . . increasingly provides a base for attaining and exercising power by the people who can claim special knowledge in matters that their society considers important.” *Id.* at 28.

See also Bayles, *supra* note 5, at 29, arguing that much of the reason experts have power over non-experts is that in a majority of the situations in which an expert opinion is needed, non-experts are experiencing some type of a crisis. “The professional has the specialized knowledge or means to diagnose the problem, determine the alternative approaches to resolving it, and then take the necessary steps. This knowledge gives a professional power over an individual.” *Id.*

166. See Gary Edmond, *Science, Law and Narrative: Helping the ‘Facts’ to Speak for Themselves*, 23 S. ILL. U. L.J. 555, 559 (1999). Edmond critiques the underpinnings of the expert witness system itself by revealing that science’s status as objective and impartial is increasingly being questioned. *Id.*; see also David S. Caudill & Richard E. Redding, *Junk Philosophy of Science?: The Paradox of Expertise and Interdisciplinarity in Federal Courts*, 57 WASH. & LEE L. REV. 685, 690 (2000) (arguing that only a weak social constructivist approach has found its way into the legal system).

167. See, e.g., BRUNO LATOUR, *SCIENCE IN ACTION: HOW TO FOLLOW SCIENTISTS AND ENGINEERS THROUGH SOCIETY* 104 (1987). Latour suggests that facts arise when readers are sufficiently convinced about something that there is no debate about it, and the processes that led to the successful persuasion of those readers has dropped out of sight. *Id.*

168. See Easton, *supra* note 65, at 471 n.15. Easton reveals, “It is axiomatic to state that a trial is a search for the truth.” *Id.*; see, e.g., *Oneida, Ltd. v. United States*, 43 Fed. Cl. 611, 619 (Fed. Cl. 1999) (stating that “the integrity and reliability of the truth finding process in a case should be paramount”); Mark R. Patterson, *Conflicts of Interest in Scientific Expert Testimony*, 40 WM & MARY L. REV. 1313 (1999). Patterson is skeptical of experts and concerned with the relationship between science and the law, but he ignores the clashes in the courtroom when experts promulgate different versions of truth. Patterson, *supra*, at 1394. Strier argues for a less adversarial and more inquisitorial judicial system. See generally Strier, *supra* note 68. He justifies his modifications, including the use of court appointed experts, by

interactions within certain social structures,¹⁶⁹ influenced by the particular needs and values of the scientists that inhabit those structures?¹⁷⁰

The conflicting guidance offered to courts by *Daubert* follows from the conflicting epistemological assumptions on which the *Daubert* decision rests. On the one hand, *Daubert* promotes the court's reliance on expertise by creating a more liberal standard in admissibility. The courts' continued emphasis on turning to experts for the "truth"¹⁷¹ demonstrates an allegiance to positivism.¹⁷² However, because the Court in *Daubert* urges judges to act as a "gatekeeper" by evaluating the relevancy and reliability of testimony, the Court is recognizing a constructivist notion:¹⁷³ the idea that truth is socially constructed.¹⁷⁴

appealing to their usefulness in reaching the truth. *Id.* at 169–70. Strier's unidimensional understanding of truth is omnipresent throughout the article. For example, "The attorney's overriding allegiance is to the client, not the truth." *Id.* at 117. And, "As opposed to the adversary system, the inquisitorial system trial is remarkably unencumbered in its search for truth." *Id.* at 144.

169. See DANIEL LITTLE, VARIETIES OF SOCIAL EXPLANATION: AN INTRODUCTION TO THE PHILOSOPHY OF SOCIAL SCIENCE 68–87 (1991). Little explains the approach to social science truth determination as the fruit of interpretative schemas derived from recognition that truth is established in relational networks.

170. See BRINT, *supra* note 4, at 145–46 (arguing that experts, while claiming to be value neutral, are actually imposing a specific set of values upon society). "[T]he social values of experts, in keeping with the main thrust of the rationalizing process, tend to favor centralized control and hierarchy, since these are associated with the virtues of predictability." *Id.* But see William Gardner et al., *Asserting Scientific Authority: Cognitive Development and Adolescent Legal Rights*, 44 AM. PSYCHOL. 895, 899 (1989) to find an example of a scientist who contends that expert scientific testimony is, in fact, value neutral. "When we claim to speak with scientific authority, it is implicit that our discourse is motivated by the epistemic values of science, rather than extra scientific values. If our scientific evaluations are filtered through our political values, why should anyone who disputes our politics accept our science?" *Id.*

171. But see Caudill & Redding, *supra* note 166, at 690. Caudill and Redding argue that courts follow a "pragmatic legal constructivis[t]" approach that ignores philosophical problems with defining science. *Id.* The authors conclude that their approach renders science in the law "almost wholly independent from the scientific enterprise." *Id.* Caudill & Redding argue against what they see as the false dichotomy of positivist approaches to science versus social constructionist approaches to science. *Id.*

172. See Margaret G. Farrell, *Daubert v. Merrell Dow Pharmaceuticals, Inc.: Epistemology and the Legal Process*, 15 CARDOZO L. REV. 2183, 2189–93 (1994) (stating that the *Frye* test's reliance on the external scientific community for determining truth is another illustration of positivism).

173. But see Caudill & Redding, *supra* note 166, at 749. In response to Browne et al.'s suggestion for court appointed experts, Caudill and Redding explain, "That is not the conclusion we would have expected from social constructivists—we were thinking something like 'court appointed experts will likely be leaders in the field who will further the hegemony of mainstream science.'" *Id.* However, whether elites would be appointed to courts as

Instead of looking to the expert for truth, the Court seems to recognize that experts offer a perspective of truth.¹⁷⁵ This Part will more carefully consider these two conflicting epistemologies.

Positivism, the epistemology¹⁷⁶ of modernism,¹⁷⁷ suggests that there is an observable reality.¹⁷⁸ Defenders of an objective science typically

experts is difficult to ascertain. The same certainly does not occur with court-appointed lawyers. Caudill and Redding have a more serious social constructivist critique of court appointed experts: "The use of court-appointed expert panels or science courts, proposed by a variety of scholars, does generally represent the view that there is a majoritarian 'correct' science. Thus, it is strongly positivist in orientation." *Id.* (citation omitted).

174. See, e.g., Michel Foucault, *Strategies of Power*, in *THE TRUTH ABOUT THE TRUTH: DECONFUSING AND RE-CONSTRUCTING THE POSTMODERN WORLD* 40, 40-45 (Walter Truett Anderson ed., 1995). Foucault wants us to focus more on identifying how what we call truth came to be known and be less concerned about what the truth is. *Id.*

175. See Anthony G. Greenwald, *The Totalitarian Ego: Fabrication and Revision of Personal History*, 35 *AM. PSYCHOL.* 603 (1980) (suggesting that humans—and thus experts—permit the external world to determine their perceptions). Specifically, Greenwald discovered humans see what is consistent with positive stories about themselves. *Id.*

176. By epistemology, we mean "How do we know what we think we know?" rather than "How do we know the *truth*?" The latter question is from a realist perspective we do not share, as will become plain.

177. Postmodernists, on the other hand, believe that there are no "truths" in the world; instead, they argue that knowledge is constructed. Consequently, an expert cannot simply report the events that occurred during an experiment. Because the expert interprets the results, she is constructing the knowledge. Consequently, postmodernists stress the importance of many voices, as opposed to a single "canonical" voice. See PAULINE MARIE ROSENAU, *POST-MODERNISM AND THE SOCIAL SCIENCES: INSIGHTS, INROADS, AND INTRUSIONS* 77-91 (1992); see also Vivienne Brown, *Decanonizing Discourses: Textual Analysis and the History of Economic Thought*, in *ECONOMICS AND LANGUAGE* 64 (Willie Henderson ed., 1993) (applying Bakhtin's process of canonization to the treatment of economic ideas and suggests that the unilogical system of the epic poem be replaced by the heteroglossia of the novel).

178. During the Enlightenment, notable philosophers including Descartes and Locke celebrated rationality and emphasized the powers of individual observation as a means of acquiring truth. See ROGER SMITH, *THE NORTON HISTORY OF THE HUMAN SCIENCES* 157, 160-83 (1997). Logical empiricists nurtured this perception and gradually the notions of "science," "rationality," and "truth" became intertwined. See Rorty, *supra* note 14, at 38-39 (discussing the impact of equating rationality with science and describing the modernist assumption that truth corresponds with reality).

In the field of psychology, behaviorism became the dominant method. Behaviorists asserted the position that through carefully controlled experiments, universal laws of human behavior like those discovered in the physical sciences would soon follow. The empiricist position that experience of the senses is the only source of knowledge, along with the principles of "physicalism" and "operationalism," legitimized behaviorists' limitations on acceptable questions, method and data, and soon the behaviorist method was considered synonymous with the scientific method. See POLKINGHORNE, *supra* note 14, at 21; see also Stephen Toulmin & David E. Leary, *The Cult of Empiricism, and Beyond*, in *A CENTURY OF PSYCHOLOGY AS SCIENCE* 594 (Sigmund Koch & David E. Leary eds., 1985).

The publication manual of the American Psychological Association provides some

focus on science itself; they argue science is “value-free,”¹⁷⁹ “honest,”

important examples of the implicit positivistic assumptions that have been incorporated into the field. See AM. PSYCHOL. ASSOC., PUBLICATION MANUAL 11 (4th ed. 1994) (demonstrating the discipline’s commitment to the view of science as a cumulative, collaborative endeavor); G. Scott Budge & Bernard Katz, *Constructing Psychological Knowledge: Reflections on Science, Scientists and Epistemology in the APA Publication Manual*, 5 THEORY & PSYCHOL. 217, 219 (1995).

This shift is more than cosmetic, for it reflects a shift from a more macroscopic to a more microscopic perspective in both form and content, as well as an implicit statement about the nature of language, science, and epistemology world-view (Weltanschauung) of psychology. These are that language is a tool, science is the accumulation of information based on single experiments that achieve positive results, knowledge is gained through the discrete and abstract measure of external behavior and recorded through the neutral and objective pen of a scientific observer, and the world is a place to be bracketed and controlled.

Budge & Katz, *supra*, at 219–20. For further discussion of modernism, see Donald N. McCloskey, *The Rhetoric of Economics*, 21 J. ECON. LIT. 481 (1983). According to McCloskey, the precepts of modernism are the following:

- (1) Prediction (and control) is the goal of science.
- (2) Only the observable implications (or predictions) of a theory matter to its truth.
- (3) Observability entails objective, reproducible experiments.
- (4) If (and only if) an experimental implication of a theory proves false is the theory proved false.
- (5) Objectivity is to be treasured; subjective “observation” (introspection) is not scientific knowledge.
- (6) Kelvin’s Dictum: “When you cannot express it in numbers, your knowledge is of a meagre and unsatisfactory kind.”
- (7) Introspection, metaphysical belief, aesthetic, and the like may well figure in the discovery of an hypothesis but cannot figure in its justification.
- (8) It is the business of methodology to demarcate scientific reasoning from non-scientific, positive from normative.
- (9) A scientific explanation of an event brings the event under a covering law.
- (10) Scientists, for instance economic scientists, have nothing to say as scientists about values, whether of morality or art.

Id. at 484.

179. See Charles Taylor, *Neutrality in Political Science*, in READINGS IN THE PHILOSOPHY OF SOCIAL SCIENCE 547–48 (Michael Martin & Lee McIntyre eds., 1994). Taylor explains the influence of positivism as a restraint on consideration of the value dimensions of research conclusions in the social sciences. The resulting effort to see research conclusions as outside the domain of values is an effort to align social science conclusions with the well-respected findings of the natural sciences.

and “free of irrational prejudice.”¹⁸⁰ They claim that science is the “ultimate market economy of knowledge, where only valid observations and plausible theories survive.”¹⁸¹ The *Frye* decision, with its reliance on “general acceptance,” is a strong example of positivism.

Social constructionism, on the other hand, offers a view of scientists as participants in a particular societal milieu, where “truth” is shaped by the various influences¹⁸² affecting the humans engaged in scientific pursuits. Constructivists argue that because social forces influence scientists, we should pay more attention to the social dimension of truth formation.¹⁸³ Thus, the Federal Rules of Evidence encourage judges and juries to consider the epistemological origins of the evidence offered by the experts, rather than simply deferring to the scientists to declare some univocal truths.

Defenders of positivism try to argue that scientists are disinterested agents who are simply recorders of reality.¹⁸⁴ In other words, they clean off the “mirror” of reality so that the rest of us can accurately see the truth.¹⁸⁵ The positivist’s reaction to social constructionism frequently

180. See Peter Atkins, *Science as Truth*, 8 HIST. HUM. SCI. 97, 97 (1995). Atkins treats science as an impersonal force, claiming that science is honest and science is free from irrational prejudice. *Id.* However, Atkins ignores the human element that influences science. See, e.g., FRED BLOCK, POSTINDUSTRIAL POSSIBILITIES: A CRITIQUE OF ECONOMIC DISCOURSE 1–32 (1990) (offering an argument against the idea of the market as an impersonal force); see also Herzberger, *supra* note 8, at 1069–72 (claiming that methodology ensures that personal viewpoints do not interfere with the unbiased collection and analysis of data; faith in “methodological procedures [is] necessary to conduct valid and reliable research”).

181. One way that the truth is uncovered is through quantification. Quantification gave credence to psychology’s claim as an exact science. See KURT DANZIGER, CONSTRUCTING THE SUBJECT: HISTORICAL ORIGINS OF PSYCHOLOGICAL RESEARCH 147 (1990) (“Quantification seemed to mark psychology as one of the exact sciences and to distinguish it sharply from such questionable pursuits as philosophy and spiritualism, with which it had been popularly associated.”).

182. See generally MICHEL FOUCAULT, POWER/KNOWLEDGE (Colin Gordon et al. trans., 1980) (arguing that when we speak we have less autonomy than we often claim because we always use the categories, argumentative strategies, metaphors, modes of composition, and rules of evidence that precede us and that have no single, identifiable author).

183. WAYNE C. BOOTH, MODERN DOGMA AND THE RHETORIC OF ASSENT xiii (1974). Booth presents the argument that truth can be found through the art of rhetoric; rhetoric being the art of discovering warrantable beliefs and improving those beliefs in shared discourse. *Id.* at 112 & n.19 (“[T]he rhetorical philosophy of Cicero” is “the fullest development of . . . consensus as a source of reliable knowledge.”).

184. See, e.g., Stephen Fuchs, *Positivism Is the Organizational Myth of Science*, 1 PERSP. ON SCI. 1, 11 (1993) (arguing that by worshipping the twin gods of scientific method and scientific community, positivism merges truth and power and presents a professional claim to cognitive monopoly as an innocent claim to truth).

185. RICHARD RORTY, PHILOSOPHY AND THE MIRROR OF NATURE 12 (1979). Rorty

casts it in the opprobrious terms of “relativism” and “nihilism”: if one believes scientific knowledge is essentially untouched by power struggles and groupthink within disciplines, then “science must be viewed as a chaotic heap of unconnected and contradictory assertions.”¹⁸⁶

However, we would argue that this overly simplistic dichotomy between absolute certitude and utter relativism is unhelpful for understanding the processes shaping expertise. Science does not operate in a vacuum; it is subject to social forces,¹⁸⁷ both within and without the scientific world. In other words, the scientists play a role in

argues against “polishing the mirror” as an educational guide. *Id.* “It is pictures rather than propositions, metaphors rather than statements, which determine most of our philosophical convictions.” *Id.*

186. Peter Huber, *Junk Science in the Courtroom*, FORBES, July 8, 1991, at 68, 72; see PAUL R. GROSS & NORMAN LEVITT, *HIGHER SUPERSTITION: THE ACADEMIC LEFT AND ITS QUARRELS WITH SCIENCE* 1–4 (1994) (arguing that leftist criticisms of science are relativistic). In reviewing this work, Fuller points out that it is the first book entirely considering threats posed by academic critics of science, but other books, namely LEWIS WOLPERT, *THE UNNATURAL NATURE OF SCIENCE* (1992), and STEVEN WEINBERG, *DREAMS OF A FINAL THEORY* (1992), addressed these criticisms in single chapters. See Steve Fuller, *A Tale of Two Cultures and Other Higher Superstitions*, 8 HIST. HUM. SCI. 115, 117 (1995).

Thomas Kuhn calls the sociology of scientific knowledge “deconstruction gone mad,” protesting that sociological interpretations of scientific processes cast unjust aspersions on the purity of scientific inquiry. THOMAS S. KUHN, *THE TROUBLE WITH THE HISTORICAL PHILOSOPHY OF SCIENCE* 8–9 (1992); see also LATOUR, *supra* note 167, at 104. An answer to Kuhn’s characterization of the sociology of scientific knowledge as an example of “deconstruction gone mad” is provided by Steven C. Ward, *In the Shadow of Deconstructed Metanarratives: Baudrillard, Latour and the End of Realist Epistemology*, 7 HIST. HUM. SCI. 73 (1994). Ward demonstrates the difference between Latour’s epistemology, based on Durkheim’s view of truth as inseparable from social organization (“truth as collective representation,” as Ward terms it), and Baudrillard’s epistemological nihilism, based on Nietzsche’s proclamation that “God [i.e. truth] is dead.” *Id.* at 75–82. “From this position, what distinguishes irrationality from rationality, belief from science, text from reality, is not the cognitive level or type of the participating actants, but the associational enhanced power of some individuals and groups to establish resistant coalitions.” *Id.* at 87.

187. See Edward Phillips, *Testing the Truth: The Alliance of Science and Law*, in CRIMINAL JUSTICE IN CRISIS 229, 239–40 (Mike McConville & Lee Bridges eds., 1994). Phillips writes,

[I]t is tempting to assume that the expert’s conclusion is objective. This is an illusion. The expert may well be operating from a theoretical or intellectual base which involves predetermined conclusions; scientific knowledge, like other forms of knowledge, does not exist in a political or institutional vacuum. This may involve methodology as well as ideology. Moreover, the conclusions drawn may involve interpretative value judgments.

Id.; see also Margaret C. Jacob, *Science and Politics in the Late Twentieth Century*, 59 SOC. RES. 487, 487–88 (1992).

creating¹⁸⁸ the image of reality they show us.¹⁸⁹

Knowledge can be legitimized; yet, we recognize that the knowledge is contested.¹⁹⁰ The relationships both within and beyond the scientific community exert influence on the creation of scientific knowledge. The next section will examine the various forces that impact science.

A. *The Birth of Scientific Fact*

When factfinders encounter science in the form of expert testimony, the expert typically testifies with phrases like, “It is firmly established.” Because this testimony is heard so often, it is important for the courts to know how such science becomes “firmly established.”

Science may be categorized as journal science or vademecum science. Journal science is frequently contradictory and consists of a diverse body of published journal articles; in contrast, vademecum science is the institutionalized knowledge of the discipline.¹⁹¹ Instead of hypothetical humility, the author uses the rhetorical language of

188. Jeremy Campbell, *Observer and Object, Reader and Text: Some Parallel Themes in Modern Science and Literature*, in *BEYOND THE TWO CULTURES: ESSAYS ON SCIENCE, TECHNOLOGY, AND LITERATURE* 23 (Joseph W. Slade & Judith Yaross Lee eds., 1990) (documenting the commonality between physics and reading in that both create understanding via observation, with special emphasis on the idea of “create”).

189. See, e.g., Daniel Bell, *The Turn to Interpretation*, 51 *PARTISAN REV.* 215, 215–19 (1984). One role scientists play in creating an image of reality is in the interpretation of their observations and “facts.” Bell argues that facts do not speak for themselves and require interpretation: “In its broadest sense, the turn to interpretation is a move away from positivism with its emphasis on naïve observation.” *Id.* at 218.

190. Joseph Rouse, *Foucault and the Natural Sciences*, in *FOUCAULT AND THE CRITIQUE OF INSTITUTIONS* 137, 158 (John Caputo & Mark Yount eds., 1993). Rouse states:

The crucial point is not that there is no legitimacy In the circulation of contested, heterogeneous knowledges, disputes about legitimacy and the criteria for legitimacy are part and parcel of the dynamics of that circulation. Understanding knowledge as “a strategical situation” rather than as a definitive outcome places epistemological reflection in the midst of ongoing struggles to legitimate (and delegitimate) various skills, practices, and assertions. Recognizing that the boundaries of science (or of knowledge) are what is being contested, epistemology is within those contested boundaries.

Id. at 158.

191. See LUDWICK FLECK, *GENESIS AND DEVELOPMENT OF SCIENTIFIC FACT* 118 (English Language ed. 1979). Fleck also refers to vademecum science as “handbook science,” a science that offers a greater degree of certainty than its counterpart. Browne et al., *supra* note 10, at 52. While journal science is contested, vademecum science is considered accepted fact; consequently, vademecum science acts as a constraint on thinking. See FLECK, *supra*, at 118–19.

reliability.

To the outsider, including judge and jury, the shift from vademecum to journal science is an unknown process, subject to metaphors¹⁹² such as “ship in a bottle” or “black box”¹⁹³—the complexities of the production facilities converting heterogeneous perspectives to homogenous scientific facts are unknown. However, Fleck argues that it is important that the court is informed of these processes.¹⁹⁴ Understanding the “genesis of a scientific fact” has important implications for the role of the expert witness.

B. *The Social Construction of Scientific Knowledge*

Robert Merton proffers four norms of science: universalism,¹⁹⁵ communism,¹⁹⁶ disinterestedness, and organized skepticism.¹⁹⁷ If these norms were a reflection of the practice of science, we could be confident about the results of scientific investigations. However, these norms are not always accurately followed; instead, science is often quite arbitrary.¹⁹⁸ The acceptance of a proposed scientific claim depends

192. See, e.g., GEORGE LAKOFF & MARK JOHNSON, *METAPHORS WE LIVE BY* (1980). The use of metaphor in scientific writing has a long and diverse history. Psychological literature is replete with metaphors describing virtually every aspect of psychological functioning. See Donal E. Carlston, *Turning Psychology on Itself: The Rhetoric of Psychology and the Psychology of Rhetoric*, in *THE RHETORIC OF THE HUMAN SCIENCES*, *supra* note 14, at 145, 146 (“[T]he theorizing of social scientists is metaphorical storytelling.”); see also James R. Averill, *Inner Feelings, Works of the Flesh, The Beast Within, Diseases of the Mind, Driving Force, and Putting on a Show: Six Metaphors of Emotion and Their Theoretical Extensions*, in *METAPHORS IN THE HISTORY OF PSYCHOLOGY* 104, 113 (David E. Leary ed., 1990) (describing metaphors of emotion including: inner feelings or experience—“He felt his anger rising,” “He ached for her”—or emotions as disease of the mind—“He was insane with rage,” “She fell madly in love”). The computer metaphor may arguably be the most familiar of cognitive metaphors. See Robert R. Hoffman et al., *Cognitive Metaphors in Experimental Psychology*, in *METAPHORS IN THE HISTORY OF PSYCHOLOGY*, *supra*, at 173, 177–89.

193. See LATOUR, *supra* note 167, at 2–3. Some scientists use a “black box” in place of a piece of machinery or a set of commands that is extremely complex. *Id.*

194. FLECK, *supra* note 191, at 121.

195. ROBERT K. MERTON, *THE SOCIOLOGY OF SCIENCE* 270–73 (1973) (according to the norm of universalism, claims would not be evaluated in a framework that considers race, gender, and status).

196. *Id.* at 275–77 (according to the norm of communism, scientists would enjoy equal accessibility to evaluative resources).

197. *Id.* at 273–75 (according to the norms of disinterestedness and organized skepticism, claims would be evaluated for their logic and validity). Furthermore, all claims receive at least the same level of scrutiny. *Id.*

198. See LATOUR, *supra* note 167, at 104. Latour states,

[T]he fate of a statement depends on others’ behaviour. You may have written the definitive paper proving that the earth is hollow or that the

largely upon those scientists¹⁹⁹ who come into contact with the proposed claim.²⁰⁰

However, the scientist is not without his own methods for gaining recognition. Scientists engage in network building in the effort to turn their claims into accepted fact. The actor-network theory of scientific development describes how they make “allies” in support of their proposed claim out of facts, people, money, theories, and organizations.²⁰¹ These actions lead to the construction of facts, as opposed to the objective discovery of facts.²⁰² Scientists compile huge reference lists creating a network of support for their position, citing even the indirect agreement obtained by referring to historical and contemporary authority, various interest groups, and informal conversations with others in their area of science.²⁰³ Those who would point to the successful scientist as being the one who develops strong networks realize that this contradicts science’s claim of pure

moon is made of green cheese but this paper will not become definitive if others do not take it up and use it as a matter of fact later on. You need *them* to make *your* paper a decisive one. If they laugh at you, if they are indifferent, if they shrug it off, that is the end of your paper.

Id. Latour restates this paragraph as a metaphor: the proposed scientific claim is a ball in a rugby game. The scientist who originally kicks the ball at one end of the field does not determine whether the ball is scored. Instead, others impact the speed and direction of the ball when they come into contact with it on the field. *Id.*

199. BENT FLYVBJERG, MAKING SOCIAL SCIENCE MATTER 32–33 (2001). Flyvbjerg argues why the social sciences are limited by what Anthony Giddens calls the “double hermeneutic.” *Id.* The people being studied are themselves making interpretations, and we must be aware of how those interpretations are formed. Then the researcher also makes interpretations. Context both determines and is determined by the researchers’ self-understanding.

200. See STEPHEN COLE, MAKING SCIENCE: BETWEEN NATURE AND SOCIETY 176 (1992) (stating that interpersonal social relations play an important role in shaping the science process).

201. See Yuval P. Yonay, *When Black Boxes Clash: Competing Ideas of What Science Is in Economics, 1924–1939*, 24 SOC. STUD. SCI. 39, 39–46 (1994) (analyzing this process at work in the struggle between neoclassical and institutionalist economists for control of the discipline and suggesting that the battle and eventual victory of the neoclassicals can be described in terms of a network-building duel, rather than rational consideration amidst Mertonian norms).

202. See generally BRUNO LATOUR & STEVE WOOLGAR, LABORATORY LIFE: THE CONSTRUCTION OF SCIENTIFIC FACTS (1986). Relying on participant observation in scientific laboratories, Latour and Woolgar argue that much lab activity attempts to create order from the disorderliness of both the world and the data collected from it. *Id.* at 246. Thus, “order,” or a claim accepted as truth, is constructed through creative processes. *Id.* at 21–23, 246.

203. Yonay, *supra* note 201, at 49–64.

objectivity.²⁰⁴ Moreover, the network-building contest is not carried out in conditions of perfect equality. The position held by an actor determines the resources and ability he has to construct networks. The community of science, like other human organizations, is stratified.²⁰⁵

C. Elites and Their Influences

One of the factors determining the status and deference awarded to individuals within the scientific community²⁰⁶ is doubtless the intellectual abilities and accomplishments of those individuals. However, this is not the only factor: the social characteristics and the operation of social processes in science also play a role in ensuring that the scientific claims of some individuals, the elites, are given attention, while other equally worthy claims may be overlooked.²⁰⁷

One way in which the elites gain an advantage is through accumulated research credit. "Once a scientist has been rewarded, his or her chance of receiving further rewards in the future are greater, *independent* of indicators of role performance," explains Cole.²⁰⁸ Those

204. Fuchs, *supra* note 184, at 4 ("[Scientists] frequently appear to support an idea not because of its sheer intellectual merit but because they have considerable investments in it.").

205. *Id.* at 12.

Sharply stratified specialties have a monopoly within a monopoly. The more one is known, the more one is heard, and so lesser known scientists working at the semi-peripheral or even marginal areas of the profession have fewer chances to attract attention to their work. Together, these processes of professional recruitment, self-referential certification, and self-reinforcing reputational inequality exclude most outsiders.

Id.

206. See WALTON, *supra* note 5, at 17, where he argues,

The main problem with the authority of science relates not so much to internal matters of how scientific reasoning is used in scientific research within the discipline to arrive at conclusions . . . as to the more subtle but more crucial problem: how the results of this research are communicated to a wider community of users who are not experts in that discipline.

See BRINT, *supra* note 4, at 203 for an interesting example of how this deification of scientific expertise within society can cause harm. "[T]he new, more exclusive emphasis on expertise . . . has primarily helped further to reduce the status of several occupations that are of great value from the perspective of their social contribution [teachers, nurses, social workers, city planners, and others]." *Id.*

207. The rhetorical force of intellectual authority has a profound impact upon the scientific audience. See generally JONATHAN R. COLE & STEPHEN COLE, *SOCIAL STRATIFICATION IN SCIENCE* (1973).

208. See COLE, *supra* note 200, at 165. Cole refers to this process as the "accumulative advantage principle," like wealth accumulation with compound interest. *Id.* Scientific

on the top of the stack tend to become more firmly established there; those on the bottom tend to be ignored.

Intellectual suppression and hegemony also influence the science that creates expert testimony.²⁰⁹ The rigidity of disciplinary boundaries limits the scientific claims that can be proffered. Disciplines will often attempt to tout their “unity”;²¹⁰ however, this highlighting often covers the diversity within the discipline and marginalizes potentially valid claims. The unity (i.e., the paradigm of the discipline) is perpetuated, not necessarily because it is valid, but because the costs of altering the paradigm are too high. Gieryn refers to a “struggle for authority, power, and resources”²¹¹ influencing scientific decision making specifically leading to the development of strongly marked boundaries between scientific and non-scientific endeavors.²¹² Gieryn argues that creating an ideology of boundary work helped establish the independence and recognition awarded to science today. Fuller corroborates this view by pointing out that disciplinary boundaries will be more clearly outlined in later accounts of the discipline’s history than in earlier ones.²¹³ Klein refers to such histories as a means of

research is certainly a worthy contribution, but this advantage means the scientist’s subsequent research is recognized over research of the non-elites, regardless of the actual worth of the research.

209. While by hegemony we refer to a Gramscian situation in which the less powerful (to their own disadvantage) accept the view of things presented by those in control of resources; intellectual suppression is the actual silencing of discordant voices. For an example of hegemony, see Jack Amariglio et al., *Division and Difference in the ‘Discipline’ of Economics*, in KNOWLEDGES: HISTORICAL AND CRITICAL STUDIES IN DISCIPLINARITY 160 (Ellen Messer-Davidow et al. eds., 1993). “[T]he portrayal of economics as a discipline with distinct boundaries is often a discursive strategy by one school or another to hegemonize the field of economic discourse.” *Id.* at 150.

210. See BROWN, *supra* note 177, at 67–68 (“Bakhtin’s notion of canonization captures a discursive process by which a range of voices loses its heterogeneity or heteroglossia and becomes assimilated as a single-voiced and unified discourse.”).

211. See Thomas F. Gieryn, *Boundary-Work and the Demarcation of Science from Non-Science: Strains and Interests in Professional Ideologies of Scientists*, 48 AM. SOC. REV. 781, 792 (1983); see also Julie Thompson Klein, *Blurring, Cracking, and Crossing: Permeation and the Fracturing of Discipline*, in KNOWLEDGES: HISTORICAL AND CRITICAL STUDIES IN DISCIPLINARITY, *supra* note 209, at 185, 190 (arguing that the diversity of theories, assumptions, and methodologies within a single field is hidden by efforts to increase disciplinary recognition).

212. Gieryn, *supra* note 211, at 781–82. For example, members of the field of psychology recognized that quantitative data could be transformed into a potent source of social power for those who directed their production and interpreted their meaning to the public. See DANZIGER *supra* note 181, at 147 (“Quantitative psychological knowledge was a species of esoteric knowledge that was held to have profound social implications.”).

213. See STEVE FULLER, *SOCIAL EPISTEMOLOGY* 197–203 (1988).

“indoctrinating new entrants into a field, legitimating the field to outsiders, and controlling, promoting, or opposing change.”²¹⁴

If consolidation of resources and the formation of consensus are the dual pathways to success, what becomes of the intellectually deviant? Huber has suggested that modern scientific enlightenment makes any recurrence of the treatment accorded Galileo impossible.²¹⁵ The paradigm established by a discipline is difficult to challenge: “paradigm stickiness”²¹⁶ and “cognitive cronyism”²¹⁷ are powerful factors to be faced by the unorthodox.

Thus, we see that while the positivist relies on scientific clear-headedness, the focus on objective verification, and the ability of all scientists to have their claims fairly evaluated, there are compelling forces within science as a human endeavor²¹⁸ that expose it to the flaws of which human interaction typically partakes.²¹⁹ Expertise emerges from a process of particularism, solitariness, interestedness, and organized dogmatism²²⁰ in which the less powerful or intellectually deviant tend to be ignored.²²¹

214. See Klein, *supra* note 211, at 196.

215. See generally Peter Huber, *Medical Experts and the Ghost of Galileo*, 54 LAW & CONTEMP. PROBS., Summer 1991, at 119. He assures us that the Galileo problem, still a concern for modern observers of science, belonged only to the unenlightened past of “scientific prehistory.” *Id.* at 168. Nowadays, the true “cranks” are obvious. *Id.* at 169.

216. See Mark A. Zupan, *Paradigms and Cultures: Some Economic Reasons for Their Stickiness*, 50 AM. J. ECON. & SOC. 99 (1991). The reasons for this “stickiness” include the fear of the established scientist that alterations in the reigning paradigm would lessen the worth of their own work, the start-up costs of a new paradigm and the costs already sunk in building the old one, the scarcity of scientists willing to risk their positions in the current paradigm by standing up for a new idea, and the poor quality of communication between paradigms. *Id.* at 99–103.

217. See G.D.L. Travis & H.M. Collins, *New Light on Old Boys: Cognitive and Institutional Particularism in the Peer Review System*, 16 SCI. TECH. & HUM. VALUES 322 (1991). Travis and Collins, who had access to ten United Kingdom Science and Engineering Research Council meetings, noted “that committee members sometimes make decisions [to award grants] based on their membership in scientific schools of thought.” *Id.* at 323. This expression of solidarity is termed by them “cognitive cronyism”: the evaluator favors the evaluatee who is similar to himself. *Id.* at 327.

218. See Henry David Thoreau, *Civil Disobedience*, in HENRY DAVID THOREAU, WALDEN: OR LIFE IN THE WOODS; ON THE DUTY OF CIVIL DISOBEDIENCE 281, 291 (Holt, Rinehart & Winston 16th ed., 1964) (“For it is, after all, with men and not with parchment that I quarrel.”).

219. See Lawrence J. Prelli, *The Rhetorical Construction of Scientific Ethos*, in RHETORIC IN THE HUMAN SCIENCES, *supra* note 16, at 48, 48–68.

220. These concepts are counter-norms of Merton’s four: universalism, communism, disinterestedness, and organized skepticism. See *supra* notes 195–197.

221. R.G.A. Dolby, *Reflections on Deviant Science*, in ON THE MARGINS OF SCIENCE: THE SOCIAL CONSTRUCTION OF REJECTED KNOWLEDGE 9 (Ron Wallis ed., 1979). The

Perhaps the best way to show the situated nature of expert testimony is to examine one particular form of expert advice under a microscope. Courts have various reasons to know when a particular party is or is not dangerous. Not too surprisingly, they often wish for experts to come forward to assure the factfinder that the truly dangerous have been identified, while those who only appear dangerous are liberated from that often damaging designation. But if what we have suggested in the first half of this Article is true, we can expect that courts relying on such expertise, unless guided by expertise from multiple legitimate perspectives on this matter, will risk significant error.

IV. INTRODUCTION TO THE STUDY OF DANGEROUSNESS

In *Kansas v. Hendricks*,²²² the Supreme Court affirmed the right of states to legislate procedures for the civil commitment of persons who, due to a “mental abnormality” or “personality disorder,” are likely to engage in “predatory acts of sexual violence.”²²³ This decision reaffirmed the Court’s acceptance of dangerousness as an appropriate legal criterion,²²⁴ and in doing so, revived the controversy surrounding expertise and predictions of violence.²²⁵ In contrast to the Court’s

elites in science tend to agree with each other, thus making it difficult and discouraging for a deviant voice from a lower level to be heard. Deviants are often “ignored or rejected after brief criticism. A scientist who doubts whether he can make an impressive case on behalf of his deviant belief may not make the effort, and turn to other scientific matters, or even leave scientific research.” *Id.* at 17. Those who remain ambitious for change devise strategies to circumvent this conservatism. See Julia M. Allen & Lester Faigley, *Discursive Strategies for Social Change: An Alternative Rhetoric of Argument*, 14 RHETORIC REV. 142 (1995).

222. 521 U.S. 346 (1997).

223. *Id.* at 350.

224. See Christopher Slobogin, *Dangerousness and Expertise*, 133 U. PA. L. REV. 97, 103–08 (1984), for a discussion of the challenges to the legitimacy of dangerousness as a legal issue. Professor Slobogin concludes that as long as statutes specify the type of harm, the time frame, and the degree of probability that it will occur they are likely to withstand constitutional challenge. *Id.*

Dangerousness is a necessary, but not sufficient, criterion for a statute to pass constitutional review. See *Hendricks*, 521 U.S. at 358. In general, the Court requires a statute to couple proof of dangerousness with proof of some additional factor, such as mental illness or abnormality. The additional requirements “serve to limit involuntary civil confinement to those who suffer from a volitional impairment rendering them dangerous beyond their control.” *Id.*; see also *Foucha v. Louisiana*, 504 U.S. 71 (1992); *Heller v. Doe*, 509 U.S. 312 (1993).

225. See, e.g., John Kip Cornwell, *Protection and Treatment: The Permissible Civil Detention of Sexual Predators*, 53 WASH. & LEE L. REV. 1293 (1996); Sarah H. Francis, Note, *Sexually Dangerous Person Statutes: Constitutional Protections of Society and the Mentally Ill or Emotionally-Driven Punishment?*, 29 SUFFOLK U. L. REV. 125 (1995); Andrew Horwitz, *Sexual Psychopath Legislation: Is There Anywhere to Go but Backwards?*, 57 U. PITT. L. REV.

affirmative stance regarding mental health professionals' ability to predict violence,²²⁶ critics have vehemently opposed the introduction of dangerousness testimony.²²⁷

Recent cases challenging preventive detention commitments of sexual offenders provide just one example of the debate over experts' ability to predict dangerousness.²²⁸ In this portion of the Article, we

35 (1995).

226. See *Jurek v. Texas*, 428 U.S. 262, 274 (1976) (addressing the constitutionality of a portion of the Texas Penal Code that made probability of future violence a consideration in capital punishment). The petitioner argued that "it is impossible to predict future behavior and that the question is so vague as to be meaningless." *Id.* The Court rejected this argument acknowledging, "It is, of course, not easy to predict future behavior. The fact that such a determination is difficult, however, does not mean that it cannot be made. Indeed, prediction of future criminal conduct is an essential element in many of the decisions rendered throughout our criminal justice system." *Id.* at 274-75. The Court again addressed the prediction of dangerousness in *Barefoot v. Estelle*, in which the majority compared the suggestion that psychiatric testimony concerning dangerousness should be abandoned to "asking us to disinvent the wheel," asserting that "we are not persuaded that such testimony is almost entirely unreliable." 463 U.S. 880, 896, 899 (1983). Justice White stated, "We are unconvinced, however, at least as of now, that the adversary process cannot be trusted to sort out the reliable from the unreliable evidence and opinion about future dangerousness, particularly when the convicted felon has the opportunity to present his own side of the case." *Id.* at 901; see also *Shall v. Martin*, 467 U.S. 253, 278 (1984) ("[F]rom a legal point of view there is nothing inherently unattainable about a prediction of future criminal conduct."); *In re Young*, 857 P.2d 989, 1017 (Wash. 1993) (acknowledging that "prediction of dangerousness has its attendant problems" but finding violence prediction testimony "sufficiently accurate and reliable" to be admitted); *State v. Post*, 541 N.W.2d 115, 126 (Wis. 1995) ("[P]redictions of future dangerousness may be difficult, [but] they are still an attainable, in fact essential, part of our judicial process.").

227. See, e.g., *People v. Murtishaw*, 29 Cal. 3d 733, 767 (1981) (holding that psychiatric predictions of dangerousness are exceptionally unreliable and highly prejudicial to the defendant); *Barefoot*, 463 U.S. at 921 (Blackmun, J. dissenting) ("[T]he unanimous conclusion of professionals in this field [is] that psychiatric predictions of long-term future violence are wrong more often than they are right. . . . It is difficult to understand how the admission of such predictions can be justified as advancing the search for truth."); Brief for American Psychiatric Association as Amicus Curiae, *Estelle v. Smith*, 451 U.S. 454 (1981); JAY ZISKIN ET AL., *COPING WITH PSYCHIATRIC AND PSYCHOLOGICAL TESTIMONY* 12 (3rd ed., 1981); Joseph J. Cocozza & Henry J. Steadman, *The Failure of Psychiatric Predictions of Dangerousness: Clear and Convincing Evidence*, 29 RUTGERS L. REV. 1084 (1976) [hereinafter *Rutgers Study*]; Alan Dershowitz, *The Role of Psychiatry in the Sentencing Process*, 1 INT'L J.L. & PSYCHIATRY 63 (1978).

228. The revival of sexual psychopath statutes was inspired by a number of highly publicized child murder cases in which previously convicted sexual offenders were released into the community and subsequently committed another sexual offense. See, e.g., Sam Howe Verhovek, *Texas Frees Child Molester Who Warns of New Crimes*, N.Y. TIMES, Apr. 9, 1996, at B7 (quoting Larry Don McQuay, who warned that he was "doomed to eventually rape, then murder my poor little victims to keep them from telling on me"); Barry Siegel, *Locking up "Sexual Predators"*, L.A. TIMES, May 10, 1990, at A1 (describing the case of Earl Shriver, a sexual offender with a long history of violent offenses who was released from prison

propose that marked variation in the methods and reliability of expert predictions of dangerousness mandate that the judiciary take an active role in determining admissibility. Specifically, we argue that the nature of violence predictions requires judicial inquiry on a case-by-case basis to address whether proffered dangerousness testimony meets admissibility standards. We support this position on two separate grounds; the first is based on an analysis of the rules of evidence, while the second stems from an examination of variations in standards of proof.

While psychological and psychiatric dangerousness testimony²²⁹ has been introduced in a variety of contexts in both civil and criminal litigation,²³⁰ we will restrict our focus to those contexts using long-term

following the state's inability to secure civil commitment in spite of psychiatric opinion that he experienced unusually sadistic sexual fantasies because Shriner had failed to commit a "recent overt act" that indicated dangerousness under Washington's commitment statute; Shriner subsequently raped and mutilated a seven-year-old boy, abandoning him for dead).

Legislators have enacted statutes that provide for a special kind of civil commitment procedure to detain "dangerous" offenders who are unable to control their sexual behavior and thus threaten public safety. *See, e.g.*, KAN. STAT. ANN. § 59-29a01 (2005); WASH. REV. CODE ANN. § 71.09.010-.120 (West 2002 & Supp. 2008). A second response to sexual violence has focused on reducing risk to the community by enacting legislation that permits law enforcement agencies to notify residents when a sexual offender has moved into the community. As a term of conditional release, these offenders are evaluated and assigned to a risk classification that determines the comprehensiveness of the notification. This classification is based, in part, on dangerousness. *See, e.g.*, OHIO REV. CODE ANN. ch. 2950 (LexisNexis 1996).

This legislation relies on psychological or psychiatric [hereinafter "psychological"] expert testimony to prove future dangerousness.

229. *See* Slobogin, *supra* note 224, at 108 (acknowledging that while courts have occasionally relied on lay testimony to determine dangerousness, "in modern times the question of dangerousness has most often been the province of expert opinion by mental health professionals"); *cf.* *Estelle v. Smith*, 451 U.S. 454, 473 (1981) (reviewing a Texas capital sentencing statute and asserting that dangerousness "does not require resort to medical experts"; however, mental health professionals provide the vast majority of predictions of dangerousness and this portion of the Article will address this type of testimony). *See, e.g.*, *Hendricks*, 521 U.S. 346; *Barefoot*, 463 U.S. 880; *Chambers v. State*, 568 S.W.2d 313 (Tex. Crim. App. 1978).

230. *See* Saleem A. Shah, *Dangerousness: A Paradigm for Exploring Some Issues in Law and Psychology*, 33 AM. PSYCHOLOGIST 224, 225 (1978) (identifying fifteen stages of the legal process in which assessments of dangerousness are made including pretrial release hearings, juvenile transfer decisions, and civil commitment hearings). For a discussion of the evaluation of dangerousness in juveniles, see Randall T. Salekin et al., *Juvenile Waiver to Adult Criminal Courts: Prototypes for Dangerousness, Sophistication-Maturity, and Amenability to Treatment*, 7 PSYCHOL. PUB. POL'Y & L. 381 (2001). "Psychological assessment of juveniles for waiver to adult criminal courts requires systematic evaluation of dangerousness, sophistication-maturity, and amenability to treatment." *Id.* at 381. Even though these three components are required for transfers, there is still much confusion about the criteria; there are "wide differences of opinion in the understanding of the constructs held

predictions²³¹ of dangerousness to others.²³² We will focus on violence prediction testimony in three types of proceedings: capital sentencing

by lawyers, judges, and psychologists.” *Id.* at 403. To determine dangerousness, Salekin et al., report four factors: “(a) extreme unprovoked violence; (b) severe, aggressive, antisocial personality; (c) lack of remorse/guilt and empathy; and (d) leadership role in the crime.” *Id.* at 397. If these four factors are present, there is a strong case for waiver to adult court. *Id.* Salekin et al. argue that case-by-case assessment of juveniles is essential. *Id.* The Supreme Court, in *Kent v. United States*, 383 U.S. 541 (1966), defined eight criteria for determining whether a juvenile should be transferred and tried in adult court. These criteria include the following:

- (1) The seriousness of the alleged offense to the community and whether the protection of the community requires waiver.
- (2) Whether the alleged offense was committed in an aggressive, violent, premeditated or willful manner.
- (3) Whether the alleged offense was against persons or against property, greater weight being given to offenses against persons especially if personal injury resulted.
- (4) The prosecutive merit of the complaint, i.e., whether there is evidence upon which a Grand Jury may be expected to return an indictment (to be determined by consultation with the United States Attorney).
- (5) The desirability of trial and disposition of the entire offense in one court when the juvenile’s associates in the alleged offense are adults who will be charged with a crime in the U.S. District Court for the District of Columbia.
- (6) The sophistication and maturity of the juvenile as determined by consideration of his home, environmental situation, emotional attitude and pattern of living.
- (7) The record and previous history of the juvenile, including previous contacts with the Youth Aid Division, other law enforcement agencies, juvenile courts and other jurisdictions, prior periods of probation to this Court, or prior commitments to juvenile institutions.
- (8) The prospects for adequate protection of the public and the likelihood of reasonable rehabilitation of the juvenile (if he is found to have committed the alleged offense) by the use of procedures, services and facilities currently available to the Juvenile Court.

Id. at 566–67.

231. That is, those predictions that purport to make predictions of violence for one year or longer after the assessment period. We have chosen long-term prediction tasks because they are the most frequently used and most often criticized by opponents of violence prediction.

232. All fifty states provide for the civil commitment of individuals found to be “mentally disordered” or found to suffer from a “mental abnormality” and are either “dangerous to others” or present a “substantial threat to the welfare of society.” See Edward Beis, *State Involuntary Commitment Statutes*, 7 MENTAL DISABILITY L. REP. 358 (1983); Edward P. Mulvey & Charles W. Lidz, *Back to Basics: A Critical Analysis of Dangerousness Research in a New Legal Environment*, 9 LAW & HUM. BEHAV. 209 (1985).

hearings,²³³ “criminal” commitment hearings,²³⁴ and civil commitment of sexual predator hearings.²³⁵ We have selected these prediction tasks because they are representative of the range of cases in which expert testimony may be solicited.²³⁶

233. Dangerousness has been identified as a mitigating factor in a number of capital sentencing statutes. See Jeffrey L. Kirchmeier, *Aggravating and Mitigating Factors: The Paradox of Today's Arbitrary and Mandatory Capital Punishment Scheme*, 6 WM. & MARY BILL RTS. J. 345, 363 (1998) (describing the role of a defendant's “future danger” in the capital punishment scheme of the various statutes).

234. Commitment and release hearings for individuals adjudicated “not guilty by reason of insanity” (“NGRI”). In a majority of states, the standards for release are identical to those of the states’ civil commitment statutes with dangerousness being an explicit criterion for commitment. See June Resnick German & Anne C. Singer, *Punishing the Not Guilty: Hospitalization of Persons Acquitted by Reason of Insanity*, 29 RUTGERS L. REV. 1011, app. (1976).

235. The sexual psychopath statutes provide for a special class of civil commitment hearings for “dangerous” offenders unable to control their sexual behavior and thus, present a threat to the public. The evaluation of future dangerousness is a necessary component for preventive detention with mental health professionals providing this opinion evidence for the courts.

236. In each of these tasks, the standards of proof, infringement on liberty, and even definitions of dangerousness are widely variable. See Grant Morris, *Defining Dangerousness: Risking a Dangerous Definition*, 10 J. CONTEMP. LEGAL ISSUES 61 (1999). Morris argues that individual autonomy is “sacrificed on the altar of collective security” without consistency or discussion about the criteria for determining future dangerousness. *Id.* at 62–63.

The Supreme Court has not informed us what magnitude of harm, or how probable its occurrence, justifies civil commitment of a mentally disordered person as “dangerous.” Rarely have other courts considered whether the statutory definition of dangerousness assures that the civil commitment criteria reflect an appropriate balance between the individual's liberty and society's safety.

Id. at 65. Examining the ability of violence prediction testimony to meet the diverse requirements will encourage a more thorough analysis of the strengths and weaknesses of the body of research, while also modeling the type of analysis Morris proposes that members of the judiciary should undertake.

Notably absent from our discussion, however, are civil commitment hearings; although imminent prediction of dangerousness is frequent and important in emergency commitments, there are fundamental differences in the procedures and outcome of these determinations. One of the most important of these differences involves the treatment and short-term incapacitation goal of the civil commitment process that differs markedly from the criminal commitment procedures. See Slobogin, *supra* note 224, at 170–74 (discussing the differences between emergency commitment and long-term prediction contexts). Thus, we are treating such predictions as beyond the scope of the Article. Limiting analysis to this subset of prediction contexts keeps the argument within manageable boundaries while facilitating a careful examination of situations in which predictions are frequently provided and have been most frequently criticized. For a review of the civil commitment process, see, e.g., Daniel W. Shuman, *The Road to Bedlam: Evidentiary Guideposts in Civil Commitment Proceedings*, 55 NOTRE DAME LAW. 53, 54 (1979).

We begin by examining the empirical literature on the prediction of dangerousness and challenge some of the misconceptions about mental health professionals' ability to predict future violence. We provide a brief history of the "first-generation" dangerousness prediction literature and the legal community's response to this research.

V. EMPIRICAL PREDICTION STUDIES

"Junk Science,"²³⁷ "quackery,"²³⁸ and "wrong about 95% of the time"²³⁹ are just a few of the allegations that critics have leveled at experts' predictions of dangerousness.²⁴⁰ In this Part, we evaluate the credibility of these claims by exploring the research on violence prediction.

A. "First-Generation Research"

The 1970s marked the birth of the first empirical studies examining the accuracy of mental health professionals' predictions of violence in patients released from psychiatric facilities by the courts.²⁴¹ In *Baxstrom*

237. See HUBER, *supra* note 76, at 220 (coining the phrase "junk science" to describe judicial acceptance of unreliable expert testimony). With regards to predictions of dangerousness, "one could favor the death penalty and 'yet still recoil at the thought that a junk science fringe of psychiatry . . . could decide who will be sent to the gallows.'" Giannelli, *supra* note 102, at 114 (quoting HUBER, *supra* note 76, at 220) (omission in original).

238. See George E. Dix, *The Death Penalty, "Dangerousness," Psychiatric Testimony, and Professional Ethics*, 5 AM. J. CRIM. L. 151, 172 (1977) (commenting on Dr. Grigson's willingness to operate at the brink of quackery).

239. See BRUCE J. ENNIS & RICHARD D. EMERY, *THE RIGHTS OF MENTAL PATIENTS* 20 (1978).

240. Clinicians do not appear to share these concerns. In 1974, a California court imposed tort liability on mental health professionals who failed to predict the violent actions of their patients. See *Tarasoff v. Regents of Univ. of Cal.*, 529 P.2d 553 (Cal. 1974) ("*Tarasoff I*"), *reargued*, 551 P.2d 334 (Cal. 1976) ("*Tarasoff II*"). *Tarasoff I* found that mental health professionals have a duty to warn potential victims of their patients, while *Tarasoff II* revised that duty into a duty to protect. A survey of 2,875 psychiatrists, psychologists, and social workers was conducted to answer questions related to this ruling. See Daniel J. Givelber et al., *Tarasoff, Myth and Reality: An Empirical Study of Private Law in Action*, 1984 WIS. L. REV. 443, 454. The authors were particularly concerned with whether the clinicians believed that the ruling was unworkable due to the "allegedly non-existent professional standards regarding the prediction of violence." *Id.* at 453. The responding clinicians reported confidence in their ability to predict future violent acts committed by their patients, with 75% indicating they could make a prediction ranging from "probable" to "certain." *Id.* at 462.

241. A number of societal forces converged to create the opportunity to examine the predictive accuracy of predictions of violence. The civil rights movement had gained significant momentum during the early-to-mid 1960s, and these successes led civil libertarians to extend the attention beyond racial minorities to other traditionally marginalized groups including juveniles, criminal defendants, and the mentally ill. The courts followed and

v. Herold,²⁴² the Court rejected the State of New York's administrative procedure by which time expired, mentally ill inmates were confined following the expiration of their sentences. The Court concluded that the practice denied equal protection under the law.²⁴³ The significance of this case is its initiation of greater concern for the basis on which people could be determined dangerous and thus in need of mandatory commitment.²⁴⁴

Five seminal studies published between 1972 and 1980 examined the incidence of violence and criminal recidivism in patients classified as dangerous by psychiatrists but released by the courts.²⁴⁵ Taken together, the results were disappointing.

The most comprehensive and methodologically sound of the studies was performed by Harry Kozol and four other clinicians who evaluated 435 male offenders and classified them as dangerous or nondangerous prior to their community release.²⁴⁶ Clinical interviews, psychological testing, and life history information was provided from a variety of collateral sources.²⁴⁷ Clinicians classified forty-nine individuals as dangerous, seventeen of whom committed a serious assaultive act that resulted in arrest during the five-year follow-up period.²⁴⁸ Thus, clinicians' accuracy rate was 35%; almost two-thirds of individuals predicted to be dangerous were not found to have committed a violent

recognized the need to extend procedural protections to these classes of individuals. *See, e.g., In re Gault*, 387 U.S. 1 (1967) (providing juvenile offenders with constitutional privileges including the right to counsel, notice, cross-examination of witnesses, and self-incrimination); *Humphrey v. Cady*, 405 U.S. 504 (1972) (extending procedural protection to mentally ill offenders).

242. 383 U.S. 107 (1966).

243. *Id.* at 110. Chief Justice Earl Warren, writing for the majority, concluded that Baxstrom was denied equal protection under the laws by the failure of the state to provide him opportunity for jury review of his civil commitment status. *Id.*

244. The effects of this decision were to provide many more safeguards for those committed for compulsory psychiatric treatment. *See, e.g., Murel v. Baltimore City Criminal Court*, 407 U.S. 355 (1972).

245. The five studies are: (1) HENRY J. STEADMAN & JOSEPH J. COCOZZA, CAREERS OF THE CRIMINALLY INSANE (1974) [hereinafter BAXSTROM STUDY]; (2) Henry J. Steadman, *A New Look at Recidivism Among Patuxent Inmates*, 5 BULL. AM. ACAD. PSYCHIATRY & L. 200 (1977) [hereinafter *Patuxent Study*]; (3) Harry L. Kozol et al., *The Diagnosis and Treatment of Dangerousness*, 18 CRIME & DELINQ. 371 (1972) [hereinafter *Kozol study*]; (4) *Rutgers Study*, *supra* note 227; and (5) TERENCE P. THORNBERRY & JOSEPH E. JACOBY, THE CRIMINALLY INSANE: A COMMUNITY FOLLOW-UP OF MENTALLY ILL OFFENDERS (1979) [hereinafter THORNBERRY STUDY].

246. *See Kozol Study*, *supra* note 245, at 388–89.

247. *Id.* at 383.

248. *Id.* at 390.

act during the follow-up.²⁴⁹

The accuracy rates of the remaining studies support the conclusion that clinicians vastly overestimated the incidence of violence in released patients.²⁵⁰ Thus, while across studies clinicians estimated that 50% to 80% of these offenders would engage in a serious aggressive act, the actual base rate for violence in this population of individuals was found to range from 12% to 15%.²⁵¹

Response to these reports was swift and nearly universal.²⁵² Mental health and legal commentators alike rejected the contention that mental health professionals could accurately predict future acts of violence.²⁵³

249. *Id.* (reporting a false-positive rate for the study of 65%, a true-positive rate of 35%, a false negative rate of 8%, and a true negative rate of 92%).

The method of equating accuracy with false-positive rates—the group of individuals predicted to be dangerous who did not act violently—is only one of the indices appropriate for measuring the utility of a decision-making strategy. If one considers the false-negative rate—the group of individuals classified as nondangerous who went on to commit a violent act—the conclusions would likely be more optimistic. In the case of the Kozol study, false-negatives were a low 8%, indicating that clinicians were accurate in 92% of the cases in which they classified an individual as nondangerous. *Id.*; see Randy K. Otto, *On the Ability of Mental Health Professionals to “Predict Dangerousness”: A Commentary on Interpretation of the “Dangerousness” Literature*, 18 L. & PSYCHOL. REV. 43, 55–58 (1994) (describing a number of formulas for determining predictive validity). The author recommends comparing the overall rate of correct classification divided by the total number of persons classified to the overall rate of correct classification expected if subjects were predicted as a function of base rates. *Id.* See Janus & Meehl, *supra* note 11, at 47–49, for an excellent discussion of the implication of base rate on prediction problems.

250. See, e.g., BAXSTROM STUDY, *supra* note 245, at 139 (reporting a false-positive rate of 80%); JOHN MONAHAN, *THE CLINICAL PREDICTION OF VIOLENT BEHAVIOR* 44–49 (1981) (describing the accuracy of the first-generation research). In general, the false positive rates were unacceptably high, ranging from 58.7% to 86%. See *id.* at 48.

251. The incidence rate or “base rate” of violence is of great concern to clinicians and researchers alike. The ability to detect an incident is directly related to the frequency with which it occurs within the population. The optimal base rate is 50%, and as the proportion of incidents decrease, the number of false positives increase. In instances where the base rate exceeds 50%, the number of false negatives increase. See, e.g., TERRENCE W. CAMPBELL, *ASSESSING SEX OFFENDERS* 48 (2007).

252. See Monahan, *supra* note 21, at 10. “Rarely has research been so uncritically accepted and so facilely generalized by both mental health professionals and lawyers as was this first-generation research on the prediction of violence.” *Id.*

253. See, e.g., MONAHAN, *supra* note 250, at 47, 49 (declaring “psychiatrists and psychologists are accurate in no more than one out of three predictions of violent behavior over a several-year period among institutionalized populations that had both committed violence in the past (and thus had high base rates for it) and who were diagnosed as mentally ill” (emphasis omitted); Bruce J. Ennis & Thomas R. Litwack, *Psychiatry and the Presumption of Expertise: Flipping Coins in the Courtroom*, 62 CAL. L. REV. 693, 737 (1974) (emphasizing that psychiatric predictions of dangerousness are more often wrong than right); *Rutgers Study*, *supra* note 227, at 1099 (asserting that the high false positive rates provided “clear and convincing proof” of clinicians’ inability to predict dangerousness).

The courts, however, expressly rejected the conclusion that mental health professionals are incapable of predicting future dangerousness with an acceptable degree of accuracy.²⁵⁴ The apparent mandate to provide risk assessment and violence prediction led many critics to re-examine their abolitionist positions.²⁵⁵ Ultimately, pragmatists, researchers, and clinicians recognized that the courts were unwilling to exclude violence prediction testimony. Members of the discipline concluded that the best way to serve their clients was to shift the focus to improving prediction accuracy by critically analyzing the methodology of the early work.

B. Criticisms of First-Generation Literature

Methodological problems significantly affect the credibility of first generation empirical evidence. One of the most significant criticisms of these problems concerns the ways in which researchers evaluated the accuracy of outcomes.²⁵⁶ Researchers reported false positive rates ranging from 54% to 80%,²⁵⁷ leading critics to suggest that predictions of dangerousness are no less accurate than “flipping a coin.”²⁵⁸ The fatal error in the coin toss analogy involves the failure to consider the implications of base rates on prediction.²⁵⁹ Incorporating base rates into the analysis indicates that in two of the studies, clinicians’ judgments did, in fact, improve on chance.²⁶⁰

254. See *Barefoot v. Estelle*, 463 U.S. 880, 901 (1983) (implying that the legal system’s reliance on prediction of dangerousness in a variety of contexts precluded exclusion in this particular prediction task).

255. See Charles W. Lidz & Edward P. Mulvey, *Dangerousness: From Legal Definition to Theoretical Research*, 19 LAW & HUM. BEHAV. 41, 42–44 (1995) (describing Saleem Shah’s progression from abolitionist to advocate as a means of serving his forensic clients).

256. See Slobogin, *supra* note 224, at 111–14.

257. See MONAHAN, *supra* note 250, at 44–46. The Baxstrom Study acknowledged an 80% false positive rate. BAXSTROM STUDY, *supra* note 245, at 139.

258. See Ennis & Litwack, *supra* note 253, at 737 (“It is inconceivable that a judgment could be considered an ‘expert’ judgment when it is less accurate than the flip of a coin.”).

259. See Slobogin, *supra* note 224, at 111 (“In fact, knowledgeable clinicians are much better at predicting dangerousness than the random selection process suggested by the coin-flipping analogy.”); see also Albert W. Alschuler, *Preventive Pretrial Detention and the Failure of Interest-Balancing Approaches to Due Process*, 85 MICH. L. REV. 510, 539–46 (1986).

260. Using Ennis & Litwack’s coin toss method, five individuals would be predicted dangerous and five non-dangerous. See Ennis & Litwack, *supra* note 253. In the Baxstrom Study, for example, the base rate of violence was approximately two in every ten patients. BAXSTROM STUDY, *supra* note 245. If both of the truly dangerous offenders were included in the predicted dangerous group, the remaining three of five, or 60%, would be false-positives. If one of the truly dangerous fell into the predicted non-dangerous group, the false-positive rate would rise to 80% while the false-negative rate would increase to 20%. In the worst case,

Additional methodological concerns arise from definitional problems with the criterion variable. Early research studies defined dangerousness²⁶¹ as an arrest for serious violence.²⁶² Use of such a narrow, restrictive definition necessarily results in inflated false positive rates.²⁶³ The reliance upon arrest records to identify violent patients also raises methodological concerns because of the significant number of violent individuals who evade police detection.²⁶⁴ False positive rates were likely inflated by the failure of the criterion variable to detect those violent acts.

Characteristics of the sample studied,²⁶⁵ reliance on restricted

if both truly dangerous offenders fell in the non-dangerous group, the false-positive rate would equal 100%, while the false-negatives would increase to 40%. Clinicians' accuracy in the Kozol and Patuxent studies both exceeded chance levels. *Kozol Study*, *supra* note 245 (reporting an accuracy rate three times better than chance); *Patuxent Study*, *supra* note 245 (reporting an accuracy rate 1.2 times better than chance).

261. Many commentators have criticized the use of the word "dangerousness" as problematic. See Megargee, *supra* note 13, at 5 ("'Dangerousness' is an unfortunate term, for it implies there is a trait of 'dangerousness' which, like intelligence, is a relatively constant characteristic of the person being assessed. However, the degree of danger an individual represents to himself or others varies markedly as a function of a number of variables."); MONAHAN, *supra* note 250, at 4–5 ("'Dangerousness' confuses issues regarding *what* one is predicting with the *probability* one is assigning to its prediction."); see also Grisso & Appelbaum, *supra* note 13, at 623 n.3 ("Future discourse in this area might be facilitated by ridding ourselves of the phrase *predictions of dangerousness*. It has no logical meaning in the context of the behavioral and social sciences. To 'predict' is to make a statement about the likelihood of a future event or behavior. *Dangerousness* seems to refer not to an event or behavior, but to a condition that exists as a function of the presence of someone or something perceived as 'dangerous.'").

262. The definition of dangerousness in legal contexts has varied considerably, with courts generally adopting a broader definition of "dangerous behavior" in civil commitment contexts. See *Jones v. United States*, 463 U.S. 354, 364–65 (1983) ("We do not agree with petitioner's suggestion that the requisite dangerousness is not established by proof that a person committed a non-violent crime against property."). Capital sentencing statutes have articulated a more stringent definition referencing the potential to cause serious bodily injury to another person. See VA. CODE ANN. § 19.2-264.2 (2004) (requiring "a probability that defendant would commit criminal acts of violence that would constitute a continuing serious threat to society"); see also Slobogin, *supra* note 224, at 101–02.

263. See Otto, *supra* note 249, at 52.

264. Every year the crime rate exceeds the arrest and conviction rate, indicating that a significant number of offenders are not apprehended for their crimes. In violence prediction research, this could lead to a significant elevation in the number of "false-positives," when in reality prediction was more accurate than follow-up measures demonstrated.

265. The research was limited to examining patients who had been charged, convicted and had served prison terms for a prior violent offense. Additionally, the patients had been hospitalized in a secure, forensic mental facility for long periods of time. See, e.g., *Patuxent Study*, *supra* note 245 (describing the mean hospitalization period at twelve years). The degree to which patients who have not experienced long-term custodial care nor engaged in prior violent acts are different from subjects studied limits the appropriateness of generalizing

samples,²⁶⁶ the frequent use of proxy measures of future violence,²⁶⁷ the failure to consider adequate predictor variables,²⁶⁸ and the use of cross-contextual predictions of violence²⁶⁹ have all been identified as methodological deficiencies. Collectively, these methodological problems restrict both the weight and the breadth of conclusions that can be reasonably drawn from this body of research. Scholars concede that the accuracy of most kinds of violence predictions remained untested, and gradually new research programs emerged to address deficits in the early work.²⁷⁰

results to dissimilar groups.

266. The inability to include the groups of individuals predicted by psychiatrists to be violent and with whom the courts agreed was significant. By virtue of this categorization, this group remained in a secure environment where opportunity to engage in violent acts is necessarily restricted. It is likely that, had these individuals been released, some unknown proportion would have engaged in violence, thereby increasing clinicians' predictive accuracy.

267. See, e.g., THORNBERRY STUDY, *supra* note 245; *Patuxent Study*, *supra* note 245. Both of these studies relied on administrative classification to infer predictions of future violence. THORNBERRY STUDY, *supra* note 245; *Patuxent Study*, *supra* note 245. One criticism of this method is that those administrative classifications were often stale (occurring months or years prior to transfer) and thus, did not accurately represent psychiatric opinion at transfer. Additionally, it is possible that extraneous factors such as maturation or the treatment the individuals received impacted the propensity to commit violence upon release. A methodologically sound study would have used clinician ratings immediately prior to release. But see *Kozol Study*, *supra* note 245 (incorporating this method but failing to demonstrate predictive accuracy). Others have argued that psychiatrists did not, in fact, believe that the majority of patients would engage in future violence, but used characterizations of "dangerousness" to exert social control. See MONAHAN, *supra* note 250, at 50–54.

268. Although clinicians had access to a variety of information sources, analysis of decision-making strategies revealed that they relied upon only a few variables, primarily age (virtually all offenders under age fifty were classified as dangerous) and severity of index offense, in making classifications. See, e.g., *Patuxent Study*, *supra* note 245. The identification of factors or variables that are predictive of violence is an important goal of second-generation research. See, e.g., Randy K. Otto, *Prediction of Dangerous Behavior: A Review and Analysis of "Second-Generation" Research*, 5 FORENSIC REP. 103 (1992).

269. See MONAHAN, *supra* note 250, at 57–58 (discussing the inefficacy of predictions made in one context (such as a hospital) that an individual will be violent in another very different context (e.g. the community) and concluding that "cross-situational consistency of any type of behavior rarely exceeds the 'sound barrier' of a .40 correlation coefficient").

270. Professor Monahan led the field in articulating the need for improvement in violence prediction in his 1984 article calling for a second generation of scholarship and outlining a number of areas that should be emphasized when designing new research. See Monahan, *supra* note 21, at 13 (suggesting researchers consider the effect of situational variables, vary the populations under investigation, and include more short-term predictions in community settings); see also Lidz & Mulvey, *supra* note 255, at 45 ("[A] field that had seen little research activity for several years was revitalized. The issues no longer seemed settled. Instead they appeared as formidable challenges.").

C. Second-Generation Research

Initial second-generation research efforts focused on examining the base rates of violence,²⁷¹ exploring the relationship between predictor variables and violence,²⁷² and assessing violence across multiple contexts and domains.²⁷³ Virtually abandoning long-term prediction as impossible,²⁷⁴ research focused on examining mental health professionals' ability to make short-term predictions of violence.²⁷⁵

The groups of individuals studied were also expanded to include both mentally disordered and non-mentally disordered individuals,²⁷⁶

271. The overestimation of violence was the single most important error in first-generation research. Thus, information regarding the occurrence of violence was a critical issue in improving the accuracy of violence predictions. *See, e.g.*, Otto, *supra* note 268, at 104 (describing base-rate research as providing "data necessary to evaluate the predictive accuracy of professionals and information regarding the overall accuracy of dangerousness predictions").

272. Researchers expanded the variables under consideration to include static and dynamic variables that had not been attended to in first-generation research. *See infra* notes 284–90 and accompanying text.

273. Incorporating more than one type of predictor variable, as well as limiting predictions to situationally similar contexts, have been posited to reduce error and increase predictive validity. *See, e.g.*, MONAHAN, *supra* note 250, at 57–59 (suggesting that research indicates that the correlation of behavior predicted in one situation and observed in another would be low).

274. *See* Monahan, *supra* note 21, at 13 (referring to long-term predictions of violence, "There are so many nails now in that coffin that I propose we declare the issue officially dead."); Otto, *supra* note 268, at 104 n.2 ("[I]t seems well established that mental health professionals are not able to make these types of predictions with any degree of accuracy."). *But see* Douglas Mossman, *Assessing Predictions of Violence: Being Accurate About Accuracy*, 62 J. CONSULTING & CLINICAL PSYCHOL. 783, 789 (1994) (describing "an average accuracy for short-term clinical predictions . . . that is not different from the accuracy of the long-term predictions").

275. The impetus for this shift came from Monahan, *supra* note 21, at 56 (Discussing the limits of the one in three accuracy reported in first-generation research, he stated, "I believe that one situation may prove to be such an exception: prediction in short-term community contexts, such as emergency civil commitment and perhaps release on bail.").

276. *See, e.g.*, Deidre Klassen & William A. O'Connor, *Crime, Inpatient Admissions, and Violence Among Male Mental Patients*, 11 INT'L J.L. & PSYCHIATRY 305, 306–10 (1988). In this study, researchers attempted to answer questions about the relationship between hospitalization, arrest, and violence. *Id.* at 306. They concluded that the number of prior arrests, the diagnosis of substance abuse, and the number of prior admissions were all significant predictors of arrest following discharge of male mental patients (Multiple $r = .49$). *Id.* at 309. Similarly, the number of prior arrests, young age, and the number of previous admissions predicted post-release violence (Multiple $r = .26$). *Id.* This study provides some evidence that the more frequently individuals are defined as disordered, the greater the likelihood that they will also be identified as criminal (arrested), and thus, demonstrates the importance of measures to address the needs of the patient both in the mental health care system and the criminal justice system. *Id.* at 310; *see also* Jeffrey S. Janofsky et al., *Psychiatrists' Accuracy in Predicting Violent Behavior on an Inpatient Unit*, 39 HOSP. &

inmates,²⁷⁷ inpatients in psychiatric hospitals,²⁷⁸ and individuals living in the community.²⁷⁹

Criterion variables were strengthened by including re-hospitalization for violent behavior,²⁸⁰ developing methods to track patients in the

COMMUNITY PSYCHIATRY 1090 (1988); cf. Harold Hall et al., *Dangerous Myths About Predicting Dangerousness?*, 2 AM. J. FORENSIC PSYCHOL. 173, 181–84 (1984) (describing the base rates of violence in U.S. servicemen).

277. See, e.g., Rueben E. Lang et al., *Personality and Criminality in Violent Offenders*, 2 J. INTERPERSONAL VIOLENCE 179, 182–83 (1987); Jeremy F. Mills et al., *Novaco Anger Scale: Reliability and Validity Within an Adult Criminal Sample*, 5 ASSESSMENT 237, 239 (1998); Frank H. Walkey & D. Ross Gilmour, *The Relationship Between Interpersonal Distance and Violence in Imprisoned Offenders*, 11 CRIM. JUST. & BEHAV. 331, 334 (1984). The unique circumstances of the correctional environment have led social scientists to hypothesize about the relationship between environmental factors and violent behavior. A body of literature suggests that overcrowded conditions are associated with increased frequency of violence in inmates who prefer greater interpersonal distance. See, e.g., Claire Lawrence & Kathryn Andrews, *The Influence of Perceived Prison Crowding on Male Inmates' Perception of Aggressive Events*, 30 AGGRESSIVE BEHAV. 273. Results from the Walkey and Gilmour study support this hypothesis; however, the study design prevented researchers from ruling out alternative causes. Walkey & Gilmour, *supra*, at 337. When viewed together, the evidence suggests that environmental factors, such as overcrowding, are related to an increase in violent acts. *Id.* at 338.

278. See, e.g., Cathy Owen et al., *Repetitively Violent Patients in Psychiatric Units*, 49 PSYCHIATRIC SERVICES 1458 (1998) [hereinafter Owen et al., *Repetitively Violent Patients*] (reporting that the risk of inpatient violence can be significant). In this study, researchers found that a small number (12%) of patients accounted for 69% of 752 serious violent incidents. *Id.* The frequency with which inpatients engage in violent acts demonstrates the importance of including these behaviors in any outcome measure of violent acts. See Dale E. McNiel & Renée L. Binder, *Clinical Assessment of the Risk of Violence Among Psychiatric Inpatients*, 148 AM. J. PSYCHIATRY 1317 (1991); Cathy Owen et al., *Violence and Aggression in Psychiatric Units*, 49 PSYCHIATRIC SERVICES 1452 (1998) [hereinafter Owen et al., *Violence and Aggression*].

279. See Jeffrey W. Swanson et al., *Violence and Psychiatric Disorder in the Community: Evidence from the Epidemiologic Catchment Area Surveys*, 41 HOSP. & COMMUNITY PSYCHIATRY 761 (1990). This well-funded, methodologically sound study addressed the association between psychiatric disorder and assaultive behavior among patients in the community. *Id.* In this analysis, 10,000 respondents completed a structured interview designed to generate a DSM-III diagnosis. *Id.* at 761–62. This interview also addressed subject violence. *Id.* at 761. Results indicated an increased risk of violence for various psychiatric illnesses. *Id.* The highest risk was associated with those respondents with alcohol or substance abuse or dependence disorders (24.57% to 34.74%). *Id.* Although a single study rarely settles a research question, this study, with a large sample size and careful methodology, provides convincing evidence that psychiatric disorders increase the risk of violence. See also Deidre Klassen & William A. O'Connor, *Assessing the Risk of Violence in Released Mental Patients: A Cross-Validation Study*, 1 PSYCHOL. ASSESSMENT 75 (1989).

280. See Edward P. Mulvey & Charles W. Lidz, *Measuring Patient Violence in Dangerousness Research*, 17 LAW & HUM. BEHAV. 277, 278 (1993) (describing the advantages and disadvantages of using re-hospitalization as a criterion variable and concluding that it is best used as an ancillary data source); see also Owen et al., *Violence and Aggression*, *supra* note 278, at 1454.

community,²⁸¹ and using self-report measures of violence.²⁸²

Using these methodological innovations, researchers explored the relationship between future violent behavior and a number of predictor variables, which can be loosely grouped into five general categories: criminal history, dispositional, demographic and case history, clinical, and situational/contextual factors.²⁸³ In the sections that follow, we review and synthesize the research literature addressing each of these categories of variables.

1. Criminal History Variables

One relatively consistent result of the second-generation literature is that the likelihood for future violence increases with past incidents of violence.²⁸⁴ Research using both mentally disordered and non-disordered populations²⁸⁵ reports significant correlations of varying size

281. See Henry J. Steadman et al., *Designing a New Generation of Risk Assessment Research*, in VIOLENCE AND MENTAL DISORDER: DEVELOPMENTS IN RISK ASSESSMENT 247, 298–305 (John Monahan & Henry J. Steadman eds., 1994) (describing the MacArthur risk assessment project and its community follow-up procedures).

282. See, e.g., Mark R. Weinrott & Maureen Saylor, *Self-Report of Crimes Committed by Sex Offenders*, 6 J. INTERPERSONAL VIOLENCE 286, 291 (1991) (describing the discrepancy between official reports of offense history and anonymously provided self-report information). This discrepancy was marked. For the rapists in the study, official records revealed a mean of 1.8 rapes per offender and a total of 66 arrests for sexual offenses for the participants, while self-report information indicated a mean of 11.7 rapes per offender and a total of 433 actual rapes. *Id.* Those identified as child molesters showed a similar pattern of responding. *Id.* Participants in the study were individuals who had been identified as sexual predators and committed under the Washington Sexual Predator Statute, and who thus can likely be considered some of the most serious and frequent offenders whose offense patterns may not be representative of others. *Id.* This study does, however, support the contention that for the most serious offenders, use of official records may result in serious underestimation of offense history. See Gene G. Abel et al., *Behavioral Approaches to Treatment of the Violent Sex Offender*, in CLINICAL TREATMENT OF THE VIOLENT PERSON 95, 96 (Loren H. Roth ed., 1987); see also Robert Plutchik & Herman M. van Praag, *A Self-Report Measure of Violence Risk, II*, 31 COMPREHENSIVE PSYCHIATRY 450, 451 (1990) (providing a general discussion of the benefits of self-report measures of violence).

283. Many commentators and researchers have found this taxonomy useful in classifying variables. See Steadman et al., *supra* note 281, at 297 (elaborating on the distinctions of these classifications); see also Deidre Klassen & William A. O'Connor, *A Prospective Study of Predictors of Violence in Adult Male Mental Health Admissions*, 12 LAW & HUM. BEHAV. 143, 145–56 (1988).

284. The adage “the best predictor of future behavior is past behavior” appears to be supported by the research. See, e.g., MONAHAN, *supra* note 250, at 71 (“If there is one finding that overshadows all others in the area of prediction, it is that the probability of future crime increases with each prior criminal act.”). The difficulty lies in discriminating between those who may continue to commit non-violent offenses and those who will re-offend violently.

285. Although research exists for both groups, a significant majority of research on

between juvenile delinquency,²⁸⁶ number of prior arrests,²⁸⁷ prior incarcerations,²⁸⁸ previous arrest for violent crime,²⁸⁹ and self-report of violence²⁹⁰ to be predictive of future violent acts.

Severity of violence at the index offense has unexpectedly been found to be inversely correlated with future violence.²⁹¹ The counter-intuitive relationship between severity and future violence may have accounted for a number of the clinical prediction errors of first-

mentally disordered samples has failed to consider the impact of criminal history variables on violent recidivism. See, e.g., Janofsky et al., *supra* note 276; Owen et al., *Repetitively Violent Patients*, *supra* note 278. But see Klassen & O'Connor, *supra* note 276. Results from a recent meta-analysis indicate "that risk assessments of mentally disordered offenders should pay close attention to the general offender prediction literature." James Bonta et al., *The Prediction of Criminal and Violent Recidivism Among Mentally Disordered Offenders: A Meta-Analysis*, 123 PSYCHOL. BULL. 123, 137 (1998). The failure to attend to these factors may be the result of the clinicians' reluctance to address static variables whose association with risk is not modifiable with treatment.

286. See, e.g., Pamela K. Lattimore et al., *Predicting Rearrest for Violence Among Serious Youthful Offenders*, 32 J. RES. CRIME & DELINQ. 54, 76 (1995) ("The majority of these variables are significant in our analysis of youthful recidivism, suggesting that each provides additional predictive power." (describing measures of juvenile offending)); Bonta et al., *supra* note 285, at 128-34 (describing the criminal history variables generally, and juvenile delinquency variables specifically, as predictors of violent recidivism); see also David P. Farrington, *Childhood Aggression and Adult Violence: Early Precursors and Later-Life Outcomes*, in THE DEVELOPMENT AND TREATMENT OF CHILDHOOD AGGRESSION 5 (Debra J. Pepler & Kenneth H. Rubin eds., 1991); David P. Farrington, *Explaining the Beginning, Progress, and Ending of Antisocial Behavior from Birth to Adulthood*, in 3 FACTS, FRAMEWORKS, AND FORECASTS: ADVANCES IN CRIMINOLOGICAL THEORY 253 (Joan McCord ed., 1992).

287. See, e.g., Klassen & O'Connor, *supra* note 279, at 79 (reporting a significant correlation between arrest record and future violence in a cross-validation sample of .26); Grant T. Harris et al., *Violent Recidivism of Mentally Disordered Offenders: The Development of a Statistical Prediction Instrument*, 20 CRIM. JUST. & BEHAV. 315, 318-21 (1993) (describing criminal history variables as significant predictors of violent recidivism); Klassen & O'Connor, *supra* note 276 (reporting that the best predictor of subsequent arrests is the number of prior arrests).

288. See, e.g., BAXSTROM STUDY, *supra* note 245; Lattimore et al., *supra* note 286, at 64; Bonta et al., *supra* note 285.

289. See, e.g., Klassen & O'Connor, *supra* note 283, at 151 (reporting a significant effect size (.14) between arrests for violent crimes within the last year and violent recidivism); Marnie E. Rice & Grant T. Harris, *A Comparison of Criminal Recidivism Among Schizophrenic and Nonschizophrenic Offenders*, 15 INT'L J.L. & PSYCHIATRY 397 (1992).

290. See, e.g., Klassen & O'Connor, *supra* note 283, at 154; Owen et al., *Repetitively Violent Patients*, *supra* note 278, at 1460-61 (reporting that inpatient violent recidivists gave warning signs, but that their threats were not taken seriously by hospital staff); Plutchik & van Praag, *supra* note 282.

291. See QUINSEY ET AL., *supra* note 33, at 147 (describing the negative correlation between the severity of index offense and future violent recidivism as -.16); cf. Bonta et al., *supra* note 285, at 128 ("Neither a violent index offense . . . nor a sexual index offense predicted future violent behavior.").

generation research, as examination of the decision-making policies of clinicians revealed that they relied heavily on the characteristics of the index offense to predict future violence.²⁹²

Taken together, criminal history variables have consistently been correlated with violent recidivism in both mentally and non-mentally disordered offenders in virtually every study in which they have been assessed;²⁹³ however, the strength of this relationship can be described only as “small.”²⁹⁴ Thus, researchers and clinicians must consider the impact of other variables to increase the predictive accuracy of violence predictions.

2. Demographic and Case History Variables

A second group of static predictors, including sex,²⁹⁵ socioeconomic status,²⁹⁶ educational attainment,²⁹⁷ marital status,²⁹⁸ employment

292. See, e.g., BAXSTROM STUDY, *supra* note 245 (describing an analysis of the clinicians’ decision-making policies and their reliance upon the characteristics of the index crime as evidence for predictions of violence). To the extent that second-generation studies rely upon clinicians that are unaware of this relationship, they too are likely to have elevated prediction errors.

293. See *supra* notes 276–79 and accompanying text.

294. Univariate correlations are in the .15 to .25 range, while those studies reporting effect sizes identify magnitudes within the .15 range. While statistically significant, the practical significance of this relationship remains questionable. See JACOB COHEN, STATISTICAL POWER ANALYSIS FOR THE BEHAVIORAL SCIENCES 24–27 (2d ed. 1977) (proposing the following interpretation of effect size: .2 considered a small effect, .5 considered a medium effect, and .8 considered a large effect).

295. The weight of the evidence suggests that men are more likely to behave violently than women. See, e.g., Melvin S. Heller & Sandra M. Ehrlich, *Actuarial Variables in 9,600 Violent and Non-Violent Offenders Referred to a Court Psychiatric Clinic*, 4 AM. J. SOC. PSYCHIATRY 30 (1984); Dale E. McNiel et al., *Predictors of Violence in Civilly Committed Acute Psychiatric Patients*, 145 AM. J. PSYCHIATRY 965, 966 (1988). But see James E. Hastings & L. Kevin Hamberger, *Sociodemographic Predictors of Violence*, 20 PSYCHIATRIC CLINICS N. AM. 323, 327 (1997) (“[T]he bulk of the research seems to show that men have a greater tendency to behave violently; however, in some samples, especially those with serious psychiatric diagnoses, the rate of violence among women may approach that of men.”); Charles Lidz et al., *The Accuracy of Predictions of Violence to Others*, 269 JAMA 1007, 1010 (1993) (describing a sample in which violence among women was higher than among men).

296. See, e.g., David A. Pritchard, *Stable Predictors of Recidivism: A Summary*, 17 CRIMINOLOGY 15 (1979) (reporting that in a number of studies, low pre-prison income level was predictive of failure on parole); Swanson et al., *supra* note 279, at 764 (reporting that of the 10,000 subjects in the Epidemiologic Catchment Area (“ECA”) Survey, the rate of violence in the low socioeconomic status (“SES”) group was three times higher than the violence rate of the upper SES group).

297. See, e.g., Harris et al., *supra* note 287, at 318 (reporting data to suggest that on average violent recidivists completed one grade less than did non-recidivists).

298. See, e.g., *id.* (describing a significant difference between recidivists and non-

history,²⁹⁹ criminality of family members,³⁰⁰ history of childhood abuse,³⁰¹ elementary school maladjustment,³⁰² and childhood aggression,³⁰³ have been consistently correlated with future violence, although, again, the strength of the relationship is small.³⁰⁴ The relationship between violence and other demographic variables, including age³⁰⁵ and race,³⁰⁶

recidivists on “never married” variable); Wagdy Loza & Gurmeet K. Dhaliwal, *Psychometric Evaluation of the Risk Appraisal Guide (RAG): A Tool for Assessing Violent Recidivism*, 12 J. INTERPERSONAL VIOLENCE 779, 781 (1997) (describing items on a risk assessment instrument including “marital status at time of index offense”). The marital status variable has been construed as an index of individuals’ general ability to form lasting interpersonal relationships with others. Those who do not engage in these relationships tend to recidivate at higher rates than those who have.

299. See, e.g., Bonta et al., *supra* note 285, at 133 (describing a significant effect for employment problems in a meta-analysis); Harris et al., *supra* note 287, at 318 (describing employment history as a variable capable of discriminating between recidivists and non-recidivists); Klassen & O’Connor, *supra* note 283, at 77.

300. See, e.g., Harris et al., *supra* note 287; Dorothy Otnow Lewis et al., *Violent Juvenile Delinquents*, 18 J. CHILD PSYCHIATRY 307 (1979) (reporting that violent juvenile offenders reported witnessing extreme violence more frequently than did non-violent or less violent offenders); Jerome A. Yesavage et al., *Family Conflict, Psychopathology, and Dangerous Behavior by Schizophrenic Inpatients*, 8 PSYCHIATRY RES. 271 (1983) (reporting that parental fighting with someone outside the family was correlated with both violence prior to admission and inpatient violence).

301. Proponents of a modeling explanation for violence have cited evidence that children who are abused engage in an increased frequency of violent acts. See, e.g., Klassen & O’Connor, *supra* note 283, at 152 (reporting that being injured by a sibling before age fifteen was predictive of violence for both schizophrenic and non-schizophrenic male inpatients); Lewis et al., *supra* note 300, at 307 (reporting that childhood abuse differentiated between more-violent and less-violent juvenile offenders).

302. See, e.g., Deborah M. Capaldi & Gerald R. Patterson, *Can Violent Offenders Be Distinguished from Frequent Offenders: Prediction from Childhood to Adolescence*, 33 J. RES. CRIME & DELINQ. 206, 225 (1996) (“Compared with the rest of the sample, violent and nonviolent arrestees appeared to be at considerable risk at Grade 4 in family background, their own antisocial behavior, and overall adjustment.”); Loza & Dhaliwal, *supra* note 298, at 781.

303. See, e.g., LEFKOWITZ ET AL., *GROWING UP TO BE VIOLENT: A LONGITUDINAL STUDY OF THE DEVELOPMENT OF AGGRESSION* (1977); Capaldi & Patterson *supra* note 302; Lattimore et al., *supra* note 286; R. Loeber & T. Dishion, *Early Predictors of Male Delinquency: A Review*, 94 PSYCHOL. BULL. 68 (1983) (citing early youth behavior problems and aggression, later youth aggression, and antisocial behavior as consistent predictors of subsequent delinquency).

304. Univariate correlations for this class of variables have been in the .22 (“parents had physical fights with others”) to the .28 (“injured by a sibling before age 15”) range. See Klassen & O’Connor, *supra* note 283, at 151 tbl.1; see also Harris et al., *supra* note 287, at 324 (reporting univariate correlation between violent recidivism and elementary school maladjustment at .31; the highest correlations then account for 9% of the total variance).

305. A majority of the research has reported a consistent inverse relationship between age and subsequent violent offending. See, e.g., Jessica M. Tanner, “Continuing Threat” to Whom?: Risk Assessment in Virginia Capital Sentencing Hearings, 17 CAP. DEF. J. 381, 384

remains less clear.

3. Dispositional Variables

Dispositional risk factors are those individual traits or styles that are considered relatively enduring personal characteristics.³⁰⁷ This category includes risk factors such as anger,³⁰⁸ impulsivity,³⁰⁹ and psychopathy or

(2005); Swanson et al., *supra* note 279, at 764 (noting that younger people accounted for seven times as much violence as did older individuals); Loza & Dhaliwal, *supra* note 298, at 789–90 (describing violent offenders as being significantly younger at the time of the index offense than non-violent offenders); Harris et al., *supra* note 287, at 318 (reporting the average age of recidivists as six years younger than non-recidivists (23.5 to 29.7, respectively)). *But see* Robert Menzies & Christopher D. Webster, *Construction and Validation of Risk Assessments in a Six-Year Follow-up of Forensic Patients: A Tridimensional Analysis*, 63 J. CONSULTING & CLINICAL PSYCH. 766, 772 (1995) (reporting “individuals 30 years or older at time of assessment were involved in more transactions than their younger counterparts, $b = -.21$, $p < .01$ —a finding that contradicts the general expectation of an inverse correlation between age and violence propensity”); Heller & Ehrlich, *supra* note 295 (reporting that subjects who had committed multiple violent acts were significantly older than nonviolent defendants or those who had only committed only one violent crime). While Hastings and Hamberger, *supra* note 295, at 326, dismiss these findings as “anomalies,” another researcher has hypothesized that the sometimes-inconsistent findings may be due to an interaction between risk factors. *See* Dale E. McNiel, *Correlates of Violence in Psychotic Patients*, 27 PSYCHIATRIC ANNALS 683, 686–87 (1997).

306. The effect of race is another variable for which findings have frequently been contradictory. Some research has indicated a slight but significant relationship between race and violence. *See, e.g.,* Steadman et al., *supra* note 281, at 274 (suggesting that African-Americans accounted for a significantly greater number of arrests for violence than would be expected); Maureen S. Baum et al., *Predicting Violent Behavior Within a Medium Security Correctional Setting*, 3 INT’L J. ECLECTIC PSYCHOTHERAPY 18, 21 (1984) (reporting that race differentiated violent from non-violent inmates). *But see* Swanson et al., *supra* note 279, at 764 (reporting that race was unrelated to violent crime in the ECA sample when socioeconomic status was controlled for). Others have suggested that membership in a violence-accepting subculture (e.g., some gangs) may provide a better explanation than the more heterogeneous classification of race. *See* McNiel, *supra* note 305, at 687.

307. *See* VIOLENCE AND MENTAL DISORDER: DEVELOPMENTS IN RISK ASSESSMENT, *supra* note 281, at 19 (describing dispositional risk factors as “those that reflect the individual person’s predispositions, traits, tendencies, or styles”).

308. Until recently the construct of anger has been conceptualized as an everyday experience, and anger assessment procedures have been developed in primarily non-clinical populations. As a consequence, instruments were unavailable to differentiate the dysfunctional elements of anger and its relationship to violence. *See* Raymond W. Novaco, *Anger as a Risk Factor for Violence Among the Mentally Disordered*, in VIOLENCE AND MENTAL DISORDER: DEVELOPMENTS IN RISK ASSESSMENT, *supra* note 281, at 21, 21. Some research has found a relationship between anger and violence. *See, e.g.,* Thomas J. Craig, *An Epidemiologic Study of Problems Associated with Violence Among Psychiatric Inpatients*, 139 AM J. PSYCHIATRY 1262 (1982) (reporting that anger was the factor most strongly associated with patient violence in a study of 1,033 psychiatric patients); Stanley R. Kay et al., *Profiles of Aggression Among Psychiatric Patients*, 176 J. NERVOUS & MENTAL DISEASE 547 (1988) (indicating that anger was one of the strongest predictors of violence in a group of

antisocial attitudes.³¹⁰ Debate continues about the utility and

hospitalized psychiatric inpatients); *see also* Robert A. Baron, *Magnitude of Victim's Pain Cues and Level of Prior Anger Arousal as Determinants of Adult Aggressive Behavior*, 17 J. PERSONALITY & SOC. PSYCHOL. 236, 239 (1971) (reporting that angered subjects delivered more intense shocks to subjects than did non-angered subjects).

Although research suggests that anger may be predictive of violence, no published research studies have incorporated a measure of anger within a comprehensive assessment. An assessment instrument has been developed to assess the role anger plays in violence, however. *See* Mills et al., *supra* note 277, at 247 (describing the Novaco Anger Scale as “an effective measure of anger in an offender population”). The MacArthur Risk Assessment Study currently underway has included this measure in an attempt to determine the predictive validity of anger as a factor in future violent acts. *See* Henry J. Steadman et al., *Designing a New Generation of Risk Assessment Research*, in VIOLENCE AND MENTAL DISORDER: DEVELOPMENTS IN RISK ASSESSMENT, *supra* note 281, at 297. In the future, other research programs should incorporate this factor into risk assessment in an effort to improve the prediction of dangerousness.

309. Impulsiveness is “related to the control of thoughts and behavior[s].” *See* Ernest S. Barratt, *Impulsiveness and Aggression*, in VIOLENCE AND MENTAL DISORDER: DEVELOPMENTS IN RISK ASSESSMENT, *supra* note 281, at 61, 61. The failure to inhibit aggressive behaviors has been hypothesized to relate to certain types of violence. *See, e.g.*, G.V. Caprara et al., *Indicators of Impulsive Aggression: Present Status of Research on Irritability and Emotional Susceptibility Scales*, 6 PERSONALITY & INDIVIDUAL DIFFERENCES 665, 666–67 (1985); G.V. Caprara et al., *Instigation to Aggress and Escalation of Aggression Examined from a Personological Perspective: The Role of Irritability and of Emotional Susceptibility*, 9 AGGRESSIVE BEHAV. 345 (1983). Problems with assessment instruments have once again limited the assessment of this factor in violence prediction research. The MacArthur Risk Assessment Survey researchers have developed an instrument, the “Barratt Impulsiveness Scale,” that shows promise in elucidating the relationship between violence and impulsiveness. *See* Barratt, *supra*, at 63.

310. Personality disorders have been described as “an enduring pattern of inner experience and behavior that deviates markedly from the expectations of the individual’s culture, is pervasive and inflexible, has an onset in adolescence or early adulthood, is stable over time, and leads to distress or impairment.” DIAGNOSTIC AND STATISTICAL MANUAL OF MENTAL DISORDERS: DSM-IV-TR 287–88 (4th ed. 2000); Stephen D. Hart et al., *Psychopathy as a Risk Marker for Violence: Development and Validation of a Screening Version of the Revised Psychopathy Checklist*, in VIOLENCE AND MENTAL DISORDER: DEVELOPMENTS IN RISK ASSESSMENT, *supra* note 281, at 81, 81.

Interpersonally, psychopaths are grandiose, egocentric, manipulative, dominant, forceful, and coldhearted. Affectively, they display shallow and labile emotions, are unable to form long-lasting bonds to people, principles, or goals, and are lacking in empathy, anxiety, and genuine guilt or remorse. Behaviorally, psychopaths are impulsive and sensation-seeking, and tend to violate social norms; the most obvious expressions of these predispositions involve criminality, substance abuse, and a failure to fulfill social obligations and responsibilities.

Id. The Psychopathy Checklist (“PCL”) (and subsequent revisions) was developed to assess psychopathy, and reliability and validity of the measure appears to be good. *See, e.g.*, Robert D. Hare et al., *Psychopathy and the DSM-IV Criteria for Antisocial Personality Disorder*, 100 J. ABNORMAL PSYCHOL. 391 (1991); Robert D. Hare et al., *The Revised Psychopathy Checklist: Reliability and Factor Structure*, 2 PSYCHOL. ASSESSMENT 338 (1990).

meaningfulness of these factors, but generally findings suggest that they have some predictive validity within limited populations.³¹¹

4. Clinical Variables

This category encompasses the relationship between diagnosis or symptomology of various types of mental disorders and violence. The impact of diagnostic categories including schizophrenia³¹² and mood

Subsequent research has demonstrated a relationship between psychopathy and violence. See, e.g., QUINSEY ET AL., *supra* note 33, at 147 (reporting that scores on the PCL-R were the single best predictor of violent offending, obtaining a correlation of .34 with violent recidivism); Grant T. Harris et al., *Psychopathy and Violent Recidivism*, 15 LAW & HUM. BEHAV. 625, 632 (1991) (describing the rate of violent failure to be significantly higher in psychopaths at 77% compared to only 21% of nonpsychopaths); Stephen D. Hart et al., *Performance of Male Psychopaths Following Conditional Release from Prison*, 56 J. CONSULTING & CLINICAL PSYCHOL. 227 (1988) (reporting that offenders who scored in the top third of the PCL distribution were three times more likely to violate the conditions of release, and approximately four times more likely to commit a violent offense than those in the bottom third of the distribution). This research suggests that psychopathy is an important construct to consider when predicting violent recidivism; however, two recent meta-analyses reported less impressive results. Paul Gendreau et al., *Is the PCL-R Really the "Unparalleled" Measure of Offender Risk? A Lesson in Knowledge Cumulation*, 29 CRIM. JUST. & BEHAV. 397 (2002) (reporting phi coefficients of .23 and .21 for general and violent recidivism, respectively). Glenn D. Walters, *Predicting Criminal Justice Outcomes with the Psychopathy Checklist and the Lifestyle Criminality Screening Form: A Meta-Analytic Comparison*, 21 BEHAV. SCI. & L. 89 (2003) (examining only prospective studies of adult and juvenile offenders and reporting a weighted *r* of .26 for general recidivism). A subsequent study revealed that Factor 2, which assesses social deviance and antisocial behavior, appeared to be more strongly correlated with both general and violent recidivism (*rw* = .32 and .26, respectively) than did Factor 1, which taps the affective and interpersonal components of psychopathy (e.g., callousness and superficial charm). See Glenn D. Walters, *Predicting Institutional Adjustment and Recidivism with the Psychopathy Checklist Factor Scores: A Meta-Analysis*, 27 LAW & HUM. BEHAV. 541, 541–51 (2003).

311. Additional research addressing the impact of these constructs on violence is an important next step in improving the accuracy of predictions. Researchers or clinicians who fail to consider the impact of dispositional factors will likely experience increased error compared to those who take a more comprehensive approach.

312. Public perception of the mentally ill as dangerous has long been the norm. In contrast, mental health professionals, motivated by the desire to reduce the stigma associated with mental illness have historically dismissed the connection between psychiatric disorder and violence. See, e.g., NAT'L MENTAL HEALTH ASSOC., *STIGMA: A LACK OF AWARENESS AND UNDERSTANDING* (1987) (pamphlet claiming that "people with mental illness pose no more of a crime threat than do other members of the general population"); MONAHAN, *supra* note 250, at 19. Recent research challenges this assertion and has demonstrated a small but consistent finding of increased risk of violence in the mentally ill when compared to the general population. See Swanson et al., *supra* note 279, at 768–69 ("The ECA data clearly demonstrate that individuals in the community with psychiatric disorders are more likely to engage in assaultive behavior . . . than those who are free of mental illness and substance abuse."); Bruce G. Link et al., *The Violent and Illegal Behavior of Mental Patients Reconsidered*, 57 AM. SOC. REV. 275, 290 (1992) ("[T]he simple assertion that mental patients

disorder,³¹³ as well as the relationship between symptoms such as delusions and hallucinations,³¹⁴ on violent recidivism has been explored. The presence of neurological impairment,³¹⁵ low intelligence,³¹⁶ and

and former mental patients are on average no more dangerous than nonpatients is incorrect.”). Currently, researchers and clinicians concede there exists a “weak association” between mental illness and community violence but stress that other factors, for example, substance abuse, are much more strongly correlated with violent acts. See John Monahan & Jean Arnold, *Violence by People with Mental Illness: A Consensus Statement by Advocates and Researchers*, 19 PSYCHIATRIC REHABILITATION J., Spring 1996, at 67, 70.

313. Affective or mood disorders include depression, mania, and bipolar disorder. Researchers have reported an increase in the prevalence of mood disturbances in violent individuals. See, e.g., Swanson et al., *supra* note 279, at 765 (“The prevalence of affective disorder was three times higher among respondents who were violent (9.37 percent) than among those who were not (2.95 percent).”); see also Renée L. Binder & Dale E. McNiel, *Effects of Diagnosis and Context on Dangerousness*, 145 AM. J. PSYCHIATRY 728 (1988); Jerome A. Yesavage, *Bipolar Illness: Correlates of Dangerous Inpatient Behaviour*, 143 BRIT. J. PSYCHIATRY 554 (1983) (reporting correlations between violence and the manic phase of bipolar illness).

314. The exploration of relationships between violence and delusional content is a relatively recent development in the literature. Delusions and hallucinations occur while individuals are in an actively psychotic state and are defined as a pathological distortion in beliefs or sensory experiences. Research suggests that delusions are the most widely experienced positive symptom of people with schizophrenia, with 90% of individuals diagnosed with schizophrenia experiencing delusions at some point during the course of their illness. See, e.g., Pamela J. Taylor et al., *Delusions and Violence*, in VIOLENCE AND MENTAL DISORDER: DEVELOPMENTS IN RISK ASSESSMENT, *supra* note 281, at 161, 165; Pamela J. Taylor et al., *Handedness and Schizophrenic Symptoms*, 55 BRIT. J. MED. PSYCHOL. 287 (1982). Preliminary research suggests an association between delusional content and violence. See, e.g., John Junginger et al., *Delusions and Symptom-Consistent Violence*, 49 PSYCHIATRIC SERV. 218, 220 (1998) (reporting results that suggest “evidence for a moderate risk that delusions would motivate violence at some time during the course of a violent subject’s illness”); see also J. Arturo Silva et al., *Delusional Misidentification and Dangerousness: A Neurobiologic Hypothesis*, 38 J. FORENSIC SCI. 904 (1993). Research efforts have been hampered, however, by the absence of a reliable and valid measure of delusional content. MacArthur Risk Assessment Survey researchers have attempted to fill this void by developing the Maudsley Assessment Delusion Schedule (“MADS”). See Taylor et al., *Delusions and Violence*, *supra*, at 178. Initial research appears promising; however, the ability of this instrument to improve the predictive accuracy of violence prediction remains unknown.

315. See, e.g., McNiel, *supra* note 305, at 684–85 (describing the association between head injury and aggressive behavior both in the first few days following the injury, and subsequently in the context of the irritability associated with post-injury personality change). He also reports data to suggest brain lesion in the temporal and orbitomedial part of the frontal lobe may elicit aggression. *Id.*; see also Menahem Krakowski, *Neurologic and Neuropsychologic Correlates of Violence*, 27 PSYCHIATRIC ANNALS 674 (1997); Menahem I. Krakowski et al., *Neurologic Impairment in Violent Schizophrenic Inpatients*, 146 AM. J. PSYCHIATRY 849 (1989).

316. See, e.g., MARTIN WOLFGANG ET AL., *DELINQUENCY IN A BIRTH COHORT* (1972) (describing the relationship between low intelligence and violence); Travis Hirschi & Michael J. Hindelang, *Intelligence and Delinquency: A Revisionist Review*, 42 AM. SOC. REV. 571

alcohol or substance abuse³¹⁷ has also been correlated with violence. Research demonstrating the relationship between clinical variables and violence has been equivocal, and generally clinical variables are considered poorer predictors of violent recidivism than are the preceding categories.³¹⁸ This may change, however, as methodological refinements in assessment instruments increase the ability to distinguish between violent and non-violent offenders.

5. Situational/Contextual Variables

A number of authors have strongly advocated for the use of situational factors to predict future violence.³¹⁹ Home environment,³²⁰

(1977) (describing the relationship between low intelligence and criminal behavior); Klassen & O'Connor, *supra* note 276 (predicting low intelligence to be related to violent recidivism).

317. The association between alcohol or substance abuse and violence is a consistent finding throughout the studies reviewed. *See, e.g.,* QUINSEY ET AL., *supra* note 33, at 147 (reporting the correlation between alcohol abuse score and violent recidivism to be .13); Richard Heyman et al., *Alcohol and Aggressive Personality Styles: Potentiators of Serious Physical Aggression Against Wives?*, 9 J. FAM. PSYCHOL. 44, 52 (1995) (discussing the finding that “husbands’ problem drinking is significantly, albeit moderately, related to serious husband-to-wife aggression in young married couples”); Andrew Johns, *Substance Misuse: A Primary Risk and a Major Problem of Comorbidity*, 9 INT’L. REV. PSYCHIATRY 233, 235 (1997) (asserting that “alcohol consumption is a major risk factor for violent offending” and that female substance abusers are equally likely to engage in violent acts as their male counterparts); Klassen & O'Connor, *supra* note 283, at 151 (reporting a significant correlation between “assaultive when drinking or using drugs” and violent recidivism of .16); Swanson et al., *supra* note 279, at 765 (“Substance abuse was by far the most prevalent diagnosis among those who were violent; 41.64 percent had alcohol or drug abuse disorders, compared with only 4.93 percent of the nonviolent respondents.”).

318. *See* Bonta et al., *supra* note 285, at 132 (reporting a mean effect size of -.03 for clinical variables, and concluding that many of the diagnostic categories were either unrelated to violence (mood disorders) or negatively related to violence (psychosis)). *But see* Susanne Strand et al., *Clinical and Risk Management Factors in Risk Prediction of Mentally Disordered Offenders—More Important Than Historical Data?*, 4 LEGAL & CRIMINOLOGICAL PSYCHOL. 67, 74 (1999) (reporting results from a Swedish study of an actuarial instrument to predict violent recidivism in which clinical and risk management factors were more predictive of violent recidivism than were historical variables).

319. *See, e.g.,* Megargee, *supra* note 13, at 6 (opining that “knowledge of both personality and situational variables was essential for accurate prediction”).

320. Research indicates that a chaotic home environment contributes to violent behavior. *See* Sue E. Estroff & Catherine Zimmer, *Social Networks, Social Support and Violence Among Persons with Severe, Persistent Mental Illness*, in VIOLENCE AND MENTAL DISORDER: DEVELOPMENTS IN RISK ASSESSMENT, *supra* note 281, at 259, 259–60 (“Chaotic, violent family environments where there was alcohol or substance abuse, an ongoing history of conflict among family members, and a controlling atmosphere were associated with violence.”); Klassen & O'Connor, *supra* note 283, at 152 (reporting significant correlations between violent recidivism and measures of dissatisfaction and conflict in the home). Homeless, mentally ill individuals are also at increased risk of violence. *See, e.g.,* McNiel, *supra* note 305, at 687 (“Previous research has indicated that the mentally ill among the

quality of interpersonal relationships,³²¹ social support,³²² treatment compliance,³²³ and presence of therapeutic alliance³²⁴ have been investigated to determine their association with future violence. Until recently, these variables had been excluded from research because they demand more frequent data collection and increasingly complex statistical analyses.³²⁵ Thus, findings regarding these variables are preliminary, and additional research is needed before a consensus concerning their impact on the predictive validity of violence prediction can be achieved.

The explosion in second-generation violence prediction research resulted in a number of philosophical and methodological problems unanticipated by researchers, scholars, or commentators.

D. Methodological Problems of Second-Generation Research

Arguably, the single most significant problem impacting this body of research is the failure of researchers to communicate or coordinate research efforts.³²⁶ Uniform definitions of predictor variables are

homeless are more likely to engage in violent behavior and felonious criminal behavior.”).

321. See, e.g., Klassen & O'Connor, *supra* note 283, at 151 (reporting significant correlations between violence and a number of interpersonal factors including “how long ago last sexual intercourse,” “dissatisfaction with siblings,” and “how long ago last relationship with a woman”).

322. See Estroff & Zimmer, *supra* note 320, at 288, 291. Estroff and Zimmer conclude that social supports are related to violence and suggest, “One way to think contextually about the risk for violence is to consider that the social network represents the opportunity to engage in violence, and social support or quality of relationships the source of provocation or perceived need for such behaviors,” but acknowledge that none of the instruments alone yielded information that was a reliable predictor in who would be violent. *Id.*

323. The impact of medication compliance on violent recidivism by psychotic patients has been established. See, e.g., Alec Buchanan & Anthony David, *Compliance and the Reduction of Dangerousness*, 3 J. MENTAL HEALTH 427 (1994); Antonio Convit et al., *Characteristics of Repeatedly Assaultive Psychiatric Inpatients*, 41 HOSP. & COMMUNITY PSYCHIATRY 1112 (1990); Jerome A. Yesavage, *Correlates of Dangerous Behavior by Schizophrenics in Hospital*, 18 J. PSYCHIATRIC RES. 225 (1984).

324. See Estroff & Zimmer, *supra* note 320, at 275 (reporting a significant bivariate relationship between involvement with mental health professionals and violence).

325. See, e.g., Mossman, *supra* note 274, at 783–86 (describing the receiver operating characteristic (“ROC”) analysis to evaluate attempts to predict violence). Asymptomatic receiver operating characteristic (“AROC”) methods describe accuracy “with indices of performance that are unaffected by base rates or by clinicians’ biases for or against Type I or Type II prediction errors.” *Id.*; see also QUINSEY ET AL., *supra* note 33, at 50–54.

326. See John Monahan & Henry J. Steadman, *Toward a Rejuvenation of Risk Assessment Research*, in VIOLENCE AND MENTAL DISORDER: DEVELOPMENTS IN RISK ASSESSMENT, *supra* note 281, at 1, 12.

This fragmentation of research efforts has seriously hindered the

virtually non-existent, follow-up periods vary considerably, and cross-validation studies are rarely undertaken.³²⁷ Inconsistent findings result in reduced confidence in both the validity and generalizability of results.

The absence of consensus concerning the appropriate measures for evaluation and communication of research findings is another impediment to the development of accuracy in violence prediction.³²⁸ Failure to recognize a standard statistic for use in analyzing and reporting results has made comparison of research findings difficult, thereby suppressing discourse and debate among mental health professionals.³²⁹

development of knowledge of the actuarial correlates of violence. The fact that each research site idiosyncratically defines its predictor and criterion variables and rarely replicates the measures used by others drastically reduces the confidence with which findings can be generalized and impedes the cumulative development of knowledge.

Id.

327. See, e.g., Norman G. Poythress, *Expert Testimony on Violence and Dangerousness: Roles for Mental Health Professionals*, 5 FORENSIC REP. 135, 143 (1992) (describing idiosyncratic predictor variables, limited cross-validation studies, and the need for larger sample, multisite research).

328. The importance of a common metric to facilitate discourse cannot be understated. An increasing number of researchers have proposed abandoning the dichotomous predictions of the past and adopting receiver operating characteristic analysis. See, e.g., QUINSEY ET AL., *supra* note 33; Douglas Mossman, *Dangerousness Decisions: An Essay on the Mathematics of Clinical Violence Prediction and Involuntary Hospitalization*, 2 U. CHI. L. SCH. ROUNDTABLE 95, 99 (1995); Mossman, *supra* note 274, at 783; cf. Stephen D. Hart et al., *A Note on Portraying the Accuracy of Violence Predictions*, 17 LAW & HUM. BEHAV. 695, 696 (1993) (describing problems with two-by-two contingency tables as a method of evaluating violence predictions).

329. An illustrative example involves the way in which researchers have defined the false-positive rate as a measure of predictive accuracy. In his influential monograph, Professor Monahan calculated the “percent false positives” in five first-generation studies. See MONAHAN, *supra* note 250, at 48. He calculated this percent by first taking all cases predicted to be violent and determined the percentage correctly predicted (true positives) and the percentage incorrectly predicted (false positives). See *id.* He then divided the sum of the false-positives by the sum of the true-positives plus the false-positives and multiplied this sum by 100, resulting in the proportion of people predicted to be violent who were *not* violent at follow-up. See *id.*

Professor Randy Otto summarized second-generation research initiated after 1981 and concluded that accuracy had improved. See Otto, *supra* note 268. He based this opinion on his calculation of false-positive error rates, a concept that is mathematically dissimilar from that used by Monahan. *Id.* Professor Otto divided the number of individuals predicted to be violent but not violent at outcome (i.e., the false positives) by the sum of the false-positives and the true-negatives—those people predicted to be non-violent who had *not* been violent at follow-up. *Id.* Monahan’s method yields higher figures than that of Otto and has led to confusion in the literature. See also Hart et al., *supra* note 328, at 697; Douglas Mossman, *Further Comments on Portraying the Accuracy of Violence Predictions*, 18 LAW & HUM.

Reliance upon weak criterion measures,³³⁰ impoverished predictor variables,³³¹ and constricted validation samples has also been identified as an ongoing methodological concern that must be addressed if substantial improvements in predictive accuracy are to occur.³³²

Methodological deficiencies notwithstanding, researchers have used information gleaned from projects examining the “cues,” or predictors of violence, to inform two very different methods of decision making.³³³ In first-generation research, clinicians who were studied relied solely upon clinical judgment to identify dangerous offenders. Proponents of this method have continued with efforts to inform clinical judgments by educating clinicians about potential variables’ predictive power and base rate estimates, then evaluating subsequent predictive efficiency of clinical judgments.³³⁴ Another contingent of investigators has combined the information gathered about individual predictors and used these cues to develop actuarial techniques³³⁵ in the effort to improve prediction accuracy. An overview of the decision-making strategies and

BEHAV. 587 (1994).

330. A number of research studies have continued to rely on “official” measures of recidivism and thus are likely underestimating violent recidivism. *See, e.g.,* Menzies & Webster, *supra* note 305, at 769 (describing reliance upon re-arrest and hospitalization data). *See generally* Mulvey & Lidz, *supra* note 280 (discussing the ways to improve methodology and criterion variable measurement).

331. The continued reluctance of researchers to include situational or contextual variables has contributed to this methodological deficiency.

332. *See* John Monahan, *Violence Prediction: The Past Twenty and the Next Twenty Years*, 23 CRIM. JUST. & BEHAV. 107, 115–17 (1996) (describing necessary advances, including “[a] rich array of theoretically chosen risk factors in multiple domains,” “[r]isk must be treated as a probability estimate that changes over time and context,” and “[l]arge and broadly representative samples of patients at multiple, coordinated sites must participate in the research”).

333. “Clinical” judgment refers to the informal, intuitive means of combining information “in one’s head,” while “actuarial” refers to a more formal, explicit decision-making strategy that combines data using a predetermined equation, table, or algorithm.

334. *See, e.g.,* William Gardner et al., *Clinical Versus Actuarial Predictions of Violence in Patients with Mental Illness*, 64 J. CONSULTING & CLINICAL PSYCHOL. 602 (1996); Janofsky et al., *supra* note 276 (indicating that psychiatrists significantly under-predicted violent behavior in a group of inpatient admissions). “We found the doctors had no ability to predict battery” Janofsky et al., *supra* note 276, at 1093. False-negative errors are unusual in clinical prediction literature and may have resulted from clinicians attempting to adjust their judgment to correspond with the relatively low base rate of violence for this population of voluntary inpatients. A small sample of only forty-seven patients also likely contributed to the error. *See id.*

335. Actuarial methods also involve gathering information in a similar way but rely on variables that have been predetermined to have some statistical correlation with future violence. These variables are then combined using a mathematical formula to produce a numerical probability that the individual will be violent during some specified time period.

their impact on accuracy follows.

VI. CLINICAL PREDICTION

Clinical judgment³³⁶ in the traditional sense involves the means by which clinicians aggregate the data they gather³³⁷ through personal interview or collateral sources concerning life history, psychometric test scores, and diagnostic impressions.³³⁸ The judgments derived from this method are often described as intuitive, and clinicians have rarely been asked to explicitly detail the processes by which they estimate risk.³³⁹ In

336. It is important to note that first-generation research studies relied solely on clinical judgment for prediction decision-making. Actuarial instruments were generated by second-generation studies that identified variables correlated with violence. It is also important to clarify that the term clinical judgment refers to the *method of aggregating data* and does not, as used by Meehl and others, “denote the judgments, inferences, observations, and practices of clinicians.” See Drew Westen & Joel Weinberger, *When Clinical Description Becomes Statistical Prediction*, 59 AM. PSYCHOLOGIST 595 (2004). A separate body of research examines the accuracy of clinicians’ observations, thought processes, and diagnostic capabilities.

337. While clinical and actuarial methods refer to the means for combining the data, a separate, but often misunderstood, issue involves the character of the data itself. Information concerning the prediction of violence has been dichotomized into “hard” variables (those data that are relatively unambiguous and can be easily scored based on verifiable information such as age, number of prior arrests, sex, or marital status) and “soft” variables (usually nonpsychometric findings that may include diagnostic impressions, clinician ratings on psychometric instruments, or other qualitative judgments). Methods that include only hard data are less susceptible to reliability challenges than those that incorporate soft factors. However, it is incorrect to assume that the decision-making strategy necessarily restricts the type of information used. Many clinical and actuarial techniques use *both* hard and soft data in predicting dangerousness. *But cf.* Slobogin, *supra* note 224, at 117–19 (describing “hard” actuarial data).

338. See PAUL MEEHL, CLINICAL VERSUS STATISTICAL PREDICTION 3–4 (1954); William M. Grove & Paul E. Meehl, *Comparative Efficiency of Informal (Subjective, Impressionistic) and Formal (Mechanical, Algorithmic) Prediction Procedures: The Clinical-Statistical Controversy*, 2 PSYCHOL. PUB. POL’Y & L. 293, 294 (1996) (“[T]he other method relies on an informal, ‘in the head,’ impressionistic, subjective conclusion, reached (somehow) by a human clinical judge.” (describing clinical judgment)); Westen & Weinberger, *supra* note 336, at 595 (“Although psychologists have revisited the question of clinical versus statistical prediction many times since Meehl’s book, . . . the weight of the evidence remains the same as it was in 1954: In the vast majority of studies, a good formula matches or trumps an intuitive clinical soothsayer . . .”).

339. See Mossman, *supra* note 328, at 100 (reporting what he considers a non-controversial assumption that the “clinical decision process that is often governed by implicit assumptions or unconscious heuristics”); Alec Buchanan, *Risk and Dangerousness*, 29 PSYCHOL. MED. 465, 466–67 (1999) (“Clinicians, on the other hand, are seldom required to describe in detail the processes by which they estimate risk.”). Empirical research suggests that we have very little awareness of factors that influence our judgments. See, e.g., Eugene F. Gauron & John K. Dickinson, *Diagnostic Decision Making in Psychiatry*, 14 ARCHIVES OF GEN. PSYCHIATRY 225 (1966); Richard E. Nisbett & Timothy D. Wilson, *Telling More Than*

this method, the focus is on the unique features of the individual for whom predictions are being made beyond the individual's membership in a broad class or classes.³⁴⁰

Historically, clinicians have pointed to their experience and training as justification for claims of accuracy concerning predictions of dangerousness; in fact, the presumption that expertise is developed through experience and training has long been accepted by the general public and the judiciary alike.³⁴¹ Nonetheless, empirical research has failed to demonstrate the relationship between experience and expertise on virtually any prediction task.³⁴² Rather, actuarial methods have consistently outperformed human judgment.³⁴³

To understand the reasons why clinical prediction and, specifically, violence determinations present such a formidable challenge, it is important to explore the impediments to accurate clinical judgment.

We Can Know: Verbal Reports on Mental Processes, 84 PSYCHOL. REV. 231 (1977).

"Heuristics" is a phrase coined by cognitive psychologists to "refer[] to implicit thinking devices that individuals use to oversimplify complex, information-processing tasks." See Mossman, *supra* note 328, at 100 n.32. Use of these procedures often leads to "systematically erroneous decisions." *Id.* For an overview of heuristics and their impact on accurate decision making, see also HOWARD N. GARB, *STUDYING THE CLINICIAN: JUDGMENT RESEARCH AND PSYCHOLOGICAL ASSESSMENT* 182–206 (1998), and *infra* note 352 and accompanying text.

340. See DAWES, *supra* note 19.

341. See, e.g., Robyn M. Dawes, *Experience and Validity of Clinical Judgments: The Illusory Correlation*, 7 BEHAV. SCI. & L. 457, 458 (1989) ("It was simply obvious to everyone involved (except me) that clinical experience was a sound basis for clinical judgment and that personal contact with people is the *sine qua non* for understanding what they are like."); see also *Jenkins v. United States*, 307 F.2d 637 (D.C. Cir. 1962).

342. See, e.g., JERRY S. WIGGINS, *PERSONALITY AND PREDICTION: PRINCIPLES OF PERSONALITY ASSESSMENT* 131–35 (1973); AM. PSYCHOLOGICAL ASSOC., *REPORT OF THE TASK FORCE ON THE EVALUATION OF EDUCATION, TRAINING AND SERVICE IN PSYCHOLOGY* (1982) (noting that there is no evidence that professional training and experience are related to professional competence); Harold N. Garb, *Clinical Judgment, Clinical Training, and Professional Experience*, 105 PSYCHOL. BULL. 387 (1989) (concluding that there is no relationship between years of clinical experience and accuracy of judgment).

343. See Grove & Meehl, *supra* note 338, at 296–300. These authors performed a meta-analysis of the existing prediction literature, and "[o]f the 136 studies [included in the analyses], 64 favored the [actuarial methods], 64 showed approximately equivalent [results], and 8 favored the clinician[']s judgments." *Id.* at 298. This discrepancy occurred in spite of the fact that a majority of the research favored the clinician by providing them with more data than used in the actuarial method. *Id.* at 299. No pattern emerged concerning the small number of studies favoring the human judges, and, in fact, fewer studies favored the clinician than would be expected by chance. See *id.* Earlier empirical analysis of this body of literature reported remarkably similar results. See also Janus & Meehl, *supra* note 11, at 48–49 ("[C]linical judgment is at best as good as, but often worse than, actuarial methods."). See generally MICHAEL WIERZBICKI, *ISSUES IN CLINICAL PSYCHOLOGY: SUBJECTIVE VERSUS OBJECTIVE APPROACHES* (1993).

Clinical prediction of violence is a complex task that involves a two-part process of identifying relevant risk factors or variables within the information provided and then assigning the appropriate weight to those factors within the prediction task.³⁴⁴ “People . . . have great difficulty combining qualitatively distinct” variables with knowledge about the distributions and the predictability of each predictor.³⁴⁵ Research into these processes has led Paul Meehl to conclude, “There are no strong arguments . . . for believing that human beings can assign optimal weights in equations subjectively or that they apply their own weights consistently”³⁴⁶

Cognitive psychologists have identified many specific impediments to the decision-making process;³⁴⁷ for example, preconceived notions or

344. See Buchanan, *supra* note 339, at 466.

345. See DAWES, *supra* note 19, at 99–100; Robyn M. Dawes, *Probabilistic Versus Causal Thinking*, in THINKING CLEARLY ABOUT PSYCHOLOGY 235, 252–61 (Dante Cicchetti & William M. Grove eds., 1991).

346. Paul E. Meehl, *Causes and Effects of My Disturbing Little Book*, 50 J. PERSONALITY ASSESSMENT 370, 372 (1986). A significant majority of academic and research psychologists have accepted Dr. Meehl’s argument, first articulated in the 1950s. See, e.g., Judith V. Becker & William D. Murphy, *What We Know and Do Not Know About Assessing and Treating Sex Offenders*, 4 PSYCHOL. PUB. POL’Y & L. 116, 124 (1998) (“We do not want to belabor the point, but, as Meehl pointed out in his classic text, actuarial prediction continues to outperform clinical prediction.”) (citations omitted); Eugene W. Wang & Pamela M. Diamond, *Empirically Identifying Factors Related to Violence Risk in Corrections*, 17 BEHAV. SCI. & L. 377, 377 (1999) (“However, ‘informal’ (clinical, subjective, impressionistic) predictions have historically been extremely poor in all domains, including aggression.”); Hart et al., *supra* note 328, at 696 (1993) (“In the context of psycholegal assessments, unwillingness to qualify one’s confidence in violence predictions or failure to make probabilistic statements regarding the likelihood of future violence is, at best, poor practice; at worst, it is simply unethical.”); Monahan & Steadman, *supra* note 326.

Nevertheless, commentators have acknowledged the failure of practitioners and clinicians to accept these assumptions concerning clinical judgment. See, e.g., CHRISTOPHER D. WEBSTER ET AL., THE HCR-20 SCHEME: THE ASSESSMENT OF DANGEROUSNESS AND RISK v (1995) (“The great challenge in what remains of the 1990’s is to integrate the almost separate worlds of research on the prediction of violence and the clinical practice of assessment. At present the two domains scarcely intersect.”); see also Randy Borum, *Improving the Clinical Practice of Violence Risk Assessment: Technology, Guidelines, and Training*, 51 AM. PSYCHOL. 945, 947 (1996) (describing the failure to incorporate research findings into a “useful, empirically based framework for clinical assessment”); Christopher D. Webster & David Cox, *Integration of Nomothetic and Ideographic Positions in Risk Assessment: Implications for Practice and the Education of Psychologists and Other Mental Health Professionals*, 52 AM. PSYCHOL. 1245, 1246 (1997) (describing the difficulties in achieving the true “scientist-practitioner”).

347. See Hal R. Arkes, *Impediments to Accurate Clinical Judgment and Possible Ways to Minimize Their Impact*, 49 J. CONSULTING & CLINICAL PSYCHOL. 323 (1981); Michael J. Saks & Robert F. Kidd, *Human Information Processing and Adjudication: Trial by Heuristics*, 15 L. & SOC’Y REV. 123, 130 (1980–1981).

expectancies have been demonstrated to impede accurate processing of information by experienced clinicians.³⁴⁸ Additional research indicates that preconceived notions also influence the perception of current data, as opposed to recalled data.³⁴⁹

Overconfidence,³⁵⁰ hindsight bias,³⁵¹ and the inability to assess covariation accurately³⁵² have all been identified as obstacles to accurate prediction. Although some recommendations concerning methods of

348. See Loren J. Chapman & Jean P. Chapman, *Genesis of Popular but Erroneous Psycho-Diagnostic Observations*, 72 J. ABNORMAL PSYCHOL. 193 (1967) (presenting drawings paired randomly with personality traits, presumably characteristic of the individual who drew the picture). Subjects fabricated illusory correlations between drawing features and personality traits (for example, large eyes were reportedly drawn by suspicious people). *Id.* at 194. Researchers proposed that the illusory correlations were influenced by a prior association between eyes and suspicion. *Id.*

349. See Arkes, *supra* note 347, at 325.

It is quite likely that all subsequent data gleaned from the client will be biased by whatever opinions have been formed during this brief initial period. Data consistent with the tentative diagnosis will be given added credence; data inconsistent with the hypothesis will be disregarded. The fact that the initial hypothesis or diagnosis is merely tentative does not decrease its biasing influence.

Id.; see also Lee Ross et al., *Social Explanation and Social Expectation: Effects of Real and Hypothetical Explanations on Subjective Likelihood*, 35 J. PERSONALITY & SOC. PSYCHOL. 817 (1977).

350. Research has consistently demonstrated the significant overconfidence clinicians have in their diagnoses. See, e.g., Stuart Oskamp, *Overconfidence in Case-Study Judgments*, 29 J. CONSULTING PSYCHOL. 261 (1965) (suggesting that providing a judge with more information increases confidence without necessarily increasing accuracy); see also James Quintar Holsopple & Joseph G. Phelan, *The Skills of Clinicians in Analysis of Projective Tests*, 10 J. CLINICAL PSYCHOL. 307, 316 (1954) (reporting that the most confident judges tend to be the least accurate).

351. This refers to the tendency of judges, when provided information about a certain outcome, to rate the probability that they would have predicted that outcome substantially higher than if they had had no information concerning the outcome. See Arkes, *supra* note 347, at 326 (describing this phenomenon); see also Baruch Fischhoff, *Hindsight ≠ Foresight: The Effect of Outcome Knowledge on Judgment Under Uncertainty*, 1 J. EXPERIMENTAL PSYCHOL.: HUM. PERCEPTION & PERFORMANCE 288 (1975).

352. See Arkes, *supra* note 347, at 323 (describing the tendency of judges who are interested in a particular symptom to only attend to those individuals who have the symptom and either do or do not develop the disease). However, in order to make an accurate determination concerning whether or not the symptom is related to the disease, the clinician must also look to the number of cases in which the symptom is absent and the disease does or does not occur. Only by considering all four groups can the hypothesis—that the symptom is predictive of the disease in this case—be adequately tested. See GARB, *supra* note 339 (providing an excellent discussion of cognitive biases, heuristics, and knowledge structures that affect the accuracy of clinical judgment, as well as identifying the representative, availability, anchoring and adjustment, and past-behavior heuristics).

rehabilitating clinical judgment have been proposed,³⁵³ a recent review of the judgment research has failed to demonstrate judgment accuracy equivalent or superior to actuarial methods.³⁵⁴

Proponents of the clinical judgment method have articulated a number of responses to critics who suggest abandoning this model.³⁵⁵ Perhaps the most frequently cited argument has become known as the “broken leg countervailings” phenomenon.³⁵⁶ Specifically, clinicians argue that, by virtue of their training and experience,³⁵⁷ they are uniquely qualified to assess the idiosyncratic factors that alter violence and subsequently incorporate these variables into the final determination of risk.³⁵⁸ Philosophical problems concerning the legitimacy of predicting ideographic outcome from nomothetic data have also been acknowledged.³⁵⁹ The convincing argument that unambiguous “facts” are rarely present in the prediction of human behavior, especially when coupled with the failure of empirical results to demonstrate improved efficacy in clinical prediction despite their ability to attend to these variables, has remained unanswered.³⁶⁰

353. Arkes, *supra* note 347, at 327–29 (suggesting that clinicians receive additional instruction in Bayesian statistics and reduce their reliance on memory, which leads to illusory correlations becoming more pronounced); *see also* David Faust & Jay Ziskin, *The Expert Witness in Psychology and Psychiatry*, 241 SCI. 31, 34 (1988).

354. *See* Grove & Meehl, *supra* note 338, at 320 (referring to clinical judgment as the “less efficient of two prediction procedures”).

355. *See, e.g.*, Hillel J. Einhorn, *Accepting Error to Make Less Error*, 50 J. PERSONALITY ASSESSMENT 387, 388–89 (1986); Thomas R. Litwack et al., *The Assessment of Dangerousness and Predictions of Violence: Recent Research and Future Prospects*, 64 PSYCHIATRIC Q. 245, 262–69 (1993); Mossman, *supra* note 274, at 783.

356. *See* MEEHL, *supra* note 338, at 25 (using an analogy to acknowledge that clinicians can detect statistically rare events—such as a broken leg—that would significantly alter the probability of an event, yet fail to be detected by an actuarial instrument). *But see* QUINSEY ET AL., *supra* note 33, at 181 (rejecting the assertion that detection of “unique psychological qualities” justifies abandonment of actuarial techniques).

357. *See supra* notes 341–42 and accompanying text.

358. DAWES, *supra* note 19, at 79 (asserting that “professional psychologists claim to be able to make predictions about individuals that transcend predictions about ‘people in general’ or about various categories of people”); Lisa Tsoi Hoshmand & Donald E. Polkinghorne, *Redefining the Science-Practice Relationship & Professional Training*, 47 AM. PSYCHOL. 55, 60 (1992) (reporting that experts’ knowledge is comprised of practice and experience that “involves accommodating previous understanding to the uniqueness of a particular clinical situation”).

359. *See, e.g.*, QUINSEY ET AL., *supra* note 33, at 180 (“People routinely use information about groups to make individual decisions.”); Grove & Meehl, *supra* note 338, at 305 (making the point that while actuarial or statistical prediction is, by nature, probabilistic rather than deterministic, statistical information can inform decision-making).

360. *See* QUINSEY ET AL., *supra* note 33, at 181 (rejecting the “broken leg countervailings” argument and providing anecdotal evidence of a case in which a serious

Others have argued that appropriate actuarial formulas do not exist for the population of offenders for which they make predictions,³⁶¹ that their skill as clinicians is greater than that of clinicians participating in the prediction research,³⁶² or that actuarial instruments have not been sufficiently evaluated to justify widespread use.³⁶³ The debate continues to rage concerning the appropriateness of clinical judgment as a prediction strategy.³⁶⁴ In spite of these concerns, reliance on clinical

physical disability—blindness—failed to incapacitate a released offender as clinicians had predicted, and the legally blind offender committed murder).

361. This argument is based on the notion that the “slight nonoptimality of beta coefficients or other statistical parameters due to validity generalization . . . would liquidate the superiority of the actuarial over the impressionistic method.” Grove & Meehl, *supra* note 338, at 301–02 (suggesting that “it does not make mathematical sense for those predictive tasks where the actuarial method’s superiority is rather strong”). Grove and Meehl conclude that if an actuarial equation predicts something with 20% greater accuracy than clinicians, and there are “no affirmative reason[s] for thinking that one’s patient group is extremely unlike all other psychiatric outpatients, . . . it is improbable . . . that a decrement of [even] 10% for the actuarial method will reduce its efficacy to the level of the clinicians.” *Id.* at 302.

Others have suggested that it is unethical to apply a prediction scheme to one’s clients without validation; however, this argument is illogical if one accepts that clinical prediction relies on invalidated anecdotal evidence in the same circumstance. *Id.* (“*Clinical experience* is only a prestigious synonym for *anecdotal evidence* when the anecdotes are told by somebody with a professional degree and a license to practice a healing art.”).

362. The underlying basis for this argument is professional arrogance. Research studies examined professional judgment from some of the most well-respected clinicians, often by those who specialized in performing these kinds of assessments, and failed to demonstrate greater accuracy than actuarial methods. Clinicians who use this argument should be challenged to produce empirical evidence of their claim or recognize that they will be discounted.

363. See *supra* notes 191–212 and accompanying text for a discussion of the validation of actuarial instruments. Additionally, the logic of this argument fails to demonstrate superiority of clinical judgment unless the clinician in question has performed some kind of outcome assessment to determine the validity of his clinical judgment; otherwise, the accuracy of both types of decision-making models remains unknown.

364. See, e.g., Kirk Heilbrun et al., *Sexual Offending: Linking Assessment, Intervention, and Decision Making*, 4 PSYCHOL. PUB. POL. & L. 138, 152 (1998) (describing the “early debate on clinical versus statistical prediction, for which research has convincingly demonstrated the superiority of actuarial prediction across a range of studies”) (citations omitted); Grove & Meehl, *supra* note 338, at 320 (“To use the less efficient of two prediction procedures in dealing with such matters is not only unscientific and irrational, it is unethical.” (referring to clinical judgment)); Grisso & Appelbaum, *supra* note 13, at 623–29 (suggesting that clinicians should avoid dichotomous, yes-no predictions of dangerousness and should instead attempt to make actuarially based risk assessments); cf. Litwack et al., *supra* note 355, at 262 (rejecting research that has questioned the legitimacy of clinical assessments of dangerousness and instead asserting that “[w]e believe that any such conclusions would be seriously misguided and would reflect a misreading of the *actual* findings—and *lack* of findings—of recent (and prior) research”); Buchanan, *supra* note 339, at 468 (“Mathematical methods have yet to provide clinicians with information of a quality to challenge the clinical judgment.”).

judgment has historically been the normative method of decision making in violence predictions.³⁶⁵

VII. ACTUARIAL PREDICTION

Actuarial or statistical methods involve the use of automatic decision rules to determine probabilities.³⁶⁶ Variables that have been statistically correlated with violence in the past are used to predict future violent acts.³⁶⁷ Information regarding the subjects' response to these variables is entered into an equation that produces a predicted likelihood of violence for that individual.³⁶⁸ Therefore, the resultant probability figure is a function of the degree of correspondence between his profile on previously identified factors and that of similar persons whose level of violence had been previously determined.³⁶⁹

Critics of the actuarial method have generally focused on its "inflexibility" or failure to consider important, case-specific information;³⁷⁰ however, consideration of these factors has failed to demonstrate substantial improvements in predictive accuracy.³⁷¹

A third type of decision making involves using actuarial prediction to inform or anchor clinical prediction.³⁷² Proponents of actuarial

365. Cf. William Gardner et al., *A Comparison of Actuarial Methods for Identifying Repetitively Violent Patients with Mental Illnesses*, 20 LAW & HUM. BEHAV. 35, 35 (1996) ("[B]ut in practice actuarial methods seem to be used rarely.").

366. See MEEHL, *supra* note 338, at 3 ("The mechanical combining of information for classification purposes, and the resultant probability figure which is an empirically determined relative frequency, are the characteristics that define the actuarial or statistical type of prediction.").

367. Regression is the most common type of equation for prediction. In this method the identified variables are first converted into a numeric form. Predictions are then made by using the regression equation to develop weighted averages of the numbers. These weights are chosen to yield the best possible prediction of future violence. See, e.g., Buchanan, *supra* note 339, at 466; DAWES, *supra* note 19, at 80–81.

368. See Stephen D. Gottfredson & Don M. Gottfredson, *Violence Prediction Methods: Statistical and Clinical Strategies*, 3 VIOLENCE & VICTIMS 303, 308–11 (1988) (describing the more commonly used statistical methods of prediction including additive linear models, clustering models, and multidimensional contingency table analysis).

369. Actuarial instruments assume that errors will be made and work to reduce that error to the smallest possible percentage. Similarly, they cannot predict whether a particular individual will commit another violent offense, but they can offer an estimate of the increased likelihood that persons with similar characteristics have engaged in violence in the past.

370. See, e.g., Buchanan, *supra* note 339, at 468; Gottfredson & Gottfredson, *supra* note 368, at 313. Note that this argument is essentially the "broken leg countervailings" rephrased. See *supra* notes 356–60 and accompanying text.

371. See generally Grove & Meehl, *supra* note 338.

372. See, e.g., MONAHAN, *supra* note 250, at 81–90 (suggesting clinicians improve predictive accuracy by using statistical data); Gottfredson & Gottfredson, *supra* note 368, at

prediction who contend that introducing unaided clinical judgment into the prediction invariably reduces predictive accuracy have rejected this “hybrid” method.³⁷³

A. Overview of the Judgment Literature

Second-generation research has attempted to demonstrate improved efficacy in clinical judgment with mixed results. Jeffrey Janofsky and his colleagues evaluated psychiatric predictions of violence in patients admitted voluntarily to a university-based inpatient unit.³⁷⁴ Results indicated that physicians demonstrated no ability to predict battery or suicide in their patients.³⁷⁵

Dale McNiel and Renée Binder assessed the accuracy of short-term clinical predictions of the potential for violence by physician and nursing staff.³⁷⁶ They concluded that staff demonstrated a moderate ability to accurately categorize patient risk but conceded that overprediction of violence continued to influence outcome.³⁷⁷

In a unique study, researchers at the University of Pittsburgh School of Medicine compared an actuarial method to clinical judgment of dangerousness.³⁷⁸ Patients presented at a psychiatric emergency room were evaluated independently by clinicians and followed for six months in the community following discharge.³⁷⁹ An actuarial instrument assessing correlates of violence was also completed using information

318 (“[P]rediction may be improved through a combined use of methods.”).

373. The ability of clinical adjustments to improve on the predictive accuracy of an actuarial prediction instrument depends upon the amount of remaining variance after the actuarial method is used, and how accurately clinicians can deal with the residual individuating information. Little evidence supports an optimistic point of view. See QUINSEY ET AL., *supra* note 33, at 65.

374. See Janofsky et al., *supra* note 276, at 1091–94. Following an intake interview that included historical and demographic data, as well as a mental status exam, psychiatrists predicted whether the patient would engage in battery or threatening or suicidal behavior over the course of the next seven days. *Id.* at 1091.

375. *Id.* at 1093. Researchers reported a slight correlation between predictions and subsequent threatening behavior, although the increased frequency of this behavior likely accounts for the correlation. *Id.*

376. See McNiel & Binder, *supra* note 278. Upon admission, nurses and physicians independently ranked the potential for violence of the patient using an overt aggression scale. *Id.* at 1319. Although the authors report a moderate level of reliability for predictions, they acknowledge that “probability estimates appeared to overpredict the rate of inpatient violence.” *Id.* at 1320.

377. *Id.* at 1320.

378. See Gardner et al., *supra* note 365.

379. Gardner et al., *supra* note 334, at 602.

from the patients' medical records, along with interview data.³⁸⁰ Statistical analysis indicated that the actuarial instruments resulted in significantly lower false-positive and false-negative rates than clinical judgment.³⁸¹

B. Development of Actuarial Instruments

The Dangerous Behavior Rating Scheme³⁸² ("DBRS") has been recognized as the first systematic attempt to develop an actuarial instrument to predict future violence.³⁸³ Based partially on a theoretical framework developed by Edwin Megargee,³⁸⁴ the semi-structured interview included personality attributes, situational factors, and perceived facilitators and inhibitors of violence.³⁸⁵ Researchers were able to achieve only modest correlations with future violence, however, and were generally pessimistic about the possibility of predicting future violence.³⁸⁶

McNiel and Binder developed a screening checklist for assessing the

380. *Id.* at 602–03.

381. *Id.* at 609. Researchers compared clinical judgment to actuarial predictions using the Receiver Operating Characteristic Curve Analysis and concluded, "Even simple actuarial models, such as those presented here, are substantially more accurate than unaided clinical prediction." *Id.*

382. See Robert Menzies et al., *The Dimensions of Dangerousness Revisited: Assessing Forensic Predictions About Violence*, 18 LAW & HUM. BEHAV. 1, 2–4 (1994).

383. See Borum, *supra* note 346, at 948–49.

384. See Megargee, *supra* note 13.

385. See Menzies et al., *supra* note 382, at 10 tbl.2 (including passive aggressive, hostility, anger, rage, emotionality, guilt, capacity for empathy, capacity for change, self-perception of dangerousness, control over actions, tolerance, emotional stress, environmental support, dangerousness increased with alcohol, dangerousness increased with drugs, and manipulative). This instrument has been considered a "hybrid" model, as it included both clinical and actuarial components. See also John Kip Cornwell, *Confining Mentally Disordered "Super Criminals": A Realignment of Rights in the Nineties*, 33 HOUS. L. REV. 651, 713 (1996).

386. Menzies et al., *supra* note 382, at 17–18 & tbl.5 (reporting correlations between DBRS and violence at one to six years ranging from .10 to .20, leading researchers to conclude, "the accuracy of both single-item and scalar predictions, as with the original study, remained decidedly unimpressive when validated against aggregate general incidents, criminal charges, and violent transactions"). The researchers went on to opine that in the 1990s, there remains "a dearth of statistically verifiable and clinically operational assessment criteria." *Id.* at 25. This failure to achieve improvements in predictive accuracy may be a result of the instruments' reliance upon clinical or impressionistic type data that have not been correlated with violence in the past. See Borum, *supra* note 346, at 948–49 (suggesting that the failure to operationalize item definitions may also have contributed to error in the instrument).

risk of violence at the time of psychiatric admission.³⁸⁷ Using a simple equally weighted linear model,³⁸⁸ they developed a five-item instrument³⁸⁹ that demonstrated some predictive accuracy in distinguishing between violent and non-violent offenders.³⁹⁰

In one of the largest, most comprehensive studies of actuarial prediction methods to date, Canadian researchers developed the Violent Risk Appraisal Guide ("VRAG").³⁹¹ Hailed by Professor Monahan as "[a] major advance in the development of actuarial risk assessment,"³⁹² the instrument was developed on a combined sample of offenders, half of whom were admitted for treatment in a maximum security psychiatric institution, and half of whom were admitted only for a brief pretrial assessment.³⁹³ Participants were 618 serious offenders,

387. See Dale E. McNiel & Renée L. Binder, *Screening for Risk of Inpatient Violence: Validation of an Actuarial Tool*, 18 LAW & HUM. BEHAV. 579 (1994).

388. See Robyn M. Dawes, *The Robust Beauty of Improper Linear Models in Decision Making*, 34 AM. PSYCHOL. 571 (1979) (contending that in many clinical situations, "improper" linear models, in which each variable is weighted equally, typically perform almost as well as "proper" linear models and require much simpler calculations). Proponents of actuarial methods have generally considered the more complex calculations an obstacle to clinical utilization of actuarial prediction. See, e.g., Gardner et al., *supra* note 365, at 35–36 (describing obstacles to the clinical use of actuarial methods). "First, clinicians may be averse to actuarial predictions because the calculations required by, for example, a regression-based method may be hard to understand and hard to perform in a clinical setting." *Id.* Thus, any method to simplify the approach may result in wider acceptance of the method.

389. By combining clinical judgment and research, the authors selected the following five items: (1) "History of physical attacks and/or fear-inducing behavior within two weeks before admission"; (2) "Absence of suicidal behavior (attempts, gestures, or threats within two weeks before admission)"; (3) "Schizophrenic or manic diagnosis"; (4) "Male gender"; and (5) "Currently married or living together." Each item was phrased so that a positive answer would increase the probability of violence (scored "one") "whereas a negative answer was scored as a 'zero.'" McNiel & Binder, *supra* note 387, at 581.

Interestingly, several of these items appear to have tapped variables whose relationship to violence remains less than clear. For example, female inpatients with serious psychiatric diagnoses have, in some studies, engaged in violence at about the same rates as their male counterparts. See Johns, *supra* note 317, at 235.

390. "When the outcome variable was limited to physical attacks, the screening checklist had a sensitivity of 55.0%, a specificity of 64.0%, a false positive rate of 67.9%, a false negative rate of 18.0%, a positive predictive value of 41.1%, a negative predictive value of 82.1%, and a total predictive value of 61.8%. The likelihood ratio was 1.52. The screening checklist had a 25.0% relative improvement over chance in classifying which patients would become physically assaultive. McNiel & Binder, *supra* note 387, at 583.

391. See QUINSEY ET AL., *supra* note 33; Harris et al., *supra* note 287.

392. See Monahan, *supra* note 332, at 113.

393. Constructing the sample in this way allows researchers some latitude in generalizing results to both forensic psychiatric patients as well as non-mentally disordered offenders who may be referred to forensic clinicians for assessment of dangerousness or pretrial assessments. See Harris et al., *supra* note 287, at 317–19.

85% of whom had been charged with at least one violent offense.³⁹⁴ The outcome or criterion variable was defined as “any new criminal charge for a violent offense.”³⁹⁵ Initially, researchers began with approximately fifty predictor variables³⁹⁶ and examined univariate correlation for each predictor.³⁹⁷ A least squares stepwise multiple regression was used to select variables that contributed independently to the prediction of violent recidivism.³⁹⁸ Twelve variables³⁹⁹ remained in the prediction equation that calculated weighted variables on their deviation from the base rate.⁴⁰⁰ Scores on the VRAG ranged from B28 to +33, with scores divided into nine groups or “bins” of eight points each. With an average follow-up of 81.5 months, the instruments’ classification accuracy was 76%.⁴⁰¹ In numerous replication studies, the VRAG continued to

394. Generalizability is limited, however, to those with serious criminal histories, and “cannot be expected to generalize to offenders with less serious criminal histories or, of course, to persons without any criminal conduct.” *Id.* at 320.

395. See QUINSEY ET AL., *supra* note 33, at 142 (defining violent offense as: “homicide, attempted homicide, kidnapping, forcible confinement, wounding, assault causing bodily harm, . . . rape,” “armed robbery,” and all sexual assaults involving physical contact).

396. Including “sociodemographic . . . (e.g., socioeconomic status, age, marital status, educational attainment, employment history); childhood problems (e.g., DSM-III conduct disorder items, [intact biological family until age sixteen], criminal history of parents and siblings, [etc.]); adult adjustment ([e.g.], psychiatric history, criminal record, . . . alcohol use, social supports . . .); characteristics of the index offense (e.g., offender’s relationship to the victim, . . . weapon used, sex of victim, motive for the index offense); and psychological assessment variables (e.g., IQ, MMPI results, Level of Supervision . . . items, . . . PCL-R score, [etc.]),” and clinical variables (e.g., expression or remorse, “insight,” etc). See *id.* at 143.

397. Variables without a significant univariate correlation were dropped from consideration. In the few pairs of variables that were highly correlated, such as a prior criminal charge and prior convictions for violent offense, the variable with the lower correlation to violent recidivism was dropped. *Id.* at 145–46.

398. *Id.* at 146 (conceding that more sophisticated statistical modeling such as event history analyses, for example, which take into account the length of time until recidivism, exist; however, application of these models found few differences in the final predictor list that was generated by the simpler method); see Marnie E. Rice & Grant T. Harris, *Cross-Validation and Extension of the Violence Risk Appraisal Guide for Child Molesters and Rapists*, 21 LAW & HUM. BEHAV. 231, 231–32 (1997).

399. The variables were: Revised Psychopathy Checklist Score (.34); elementary school maladjustment (.31); meets DSM-III criteria for any personality disorder (.26); age at the time of index offense (-.26); separation from either parent (except death) under age 16 (.25); failure on prior conditional release (.24); nonviolent offense history score (using the Cormier-Lang scale) (.20); never married or equivalent (.18); meets DSM-III criteria for schizophrenia (-.17); most serious victim injury (from index offense) (-.16); alcohol abuse score (.13); and female victim in the index offense (-.11). See QUINSEY ET AL., *supra* note 33, at 147.

400. See J. NUFFIELD, *PAROLE DECISION-MAKING IN CANADA: RESEARCH TOWARDS DECISION GUIDELINES* (1982) (discussing the method of anchoring weights to deviations from base rate response).

401. No other instrument has reported such a high degree of predictive accuracy.

perform well with no appreciable decrement in predictive validity.⁴⁰² Perhaps the most significant concern related to the VRAG is the failure of independent clinicians and researchers to demonstrate validity of the instrument; however, the instrument's accuracy has been robust, and without an "affirmative reason" for why the instrument would not apply, its application to other serious offender populations appears justified.

A second instrument, the HCR-20, is a violence risk assessment scheme that the authors contend has potential applicability to a variety of settings.⁴⁰³ The HCR-20 assesses past, present, and future indicators of risk and includes empirically derived factors to assess both static and dynamic variables associated with increased violence. The ten historical factors include items related to criminal history, demographic and case history, and individual disposition (including psychopathy) that are relatively immutable,⁴⁰⁴ while the five clinical items are intended to reflect current, dynamic correlates of violence,⁴⁰⁵ and the five risk management items address environmental, post-assessment factors that may either aggravate or mitigate risk.⁴⁰⁶ Research in a variety of settings suggests that the HCR-20 demonstrated predictive utility with a variety of groups, including male and female civil psychiatric patients,⁴⁰⁷ forensic

Additionally, the instrument enables clinicians to provide specific numerical probabilities or range of values, along with normative information about how this individual's risk score compares to other offenders. Thus, clinicians using this instrument can avoid the ethically questionable practice of providing testimony in conclusory terms, *see, e.g.*, Grisso & Appelbaum, *supra* note 13; Poythress, *supra* note 327, but can instead leave the determination of the appropriate level of risk to require intervention up to individuals elected or appointed to do so.

402. *See, e.g.*, Marnie E. Rice & Grant T. Harris, *Violent Recidivism: Assessing Predictive Validity*, 63 J. CONSULTING & CLINICAL PSYCHOL. 737 (1995); Rice & Harris, *supra* note 398 (considering this study to be a cross-validation study in that it evaluated sexual offenders not previously included in the construction sample); Vernon Quinsey et al., *Actuarial Prediction of Sexual Recidivism*, 10 J. INTERPERSONAL VIOLENCE 85 (1995). *But see* QUINSEY ET AL., *supra* note 33, at 153–55 (failing to recommend use of the VRAG with fire setters without additional empirical study).

403. *See* WEBSTER ET AL., *supra* note 346; *see also* CHRISTOPHER WEBSTER ET AL., HCR-20: ASSESSING THE RISK FOR VIOLENCE (VERSION 2) (1997).

404. The ten historical factors are: previous violence, young age at first offense, relationship instability, employment problems, substance use problems, major mental illness, psychopathy, early maladjustment, personality disorder, and prior supervision failure. WEBSTER, *supra* note 403.

405. The five clinical items are: lack of insight, negative attitudes, active symptoms of major mental illness, impulsivity, and unresponsive to treatment. *Id.*

406. The five risk management items are: plans lack feasibility, exposure to destabilizers, lack of personal support, noncompliance with remediation attempts, and stress. *Id.*

407. *See, e.g.*, Kevin S. Douglas et al., *Assessing Risk for Violence Among Psychiatric*

psychiatric patients,⁴⁰⁸ and incarcerated offenders;⁴⁰⁹ however, more recent research findings have suggested that the HCR-20's predictive accuracy diminishes in restrictive settings.⁴¹⁰ In summary, the HCR-20 has been praised for its empirical basis, operationally defined coding system, and practical use,⁴¹¹ but critics note that a majority of the research studies using the HCR-20 have been conducted abroad, and thus, questions remain about the applicability of findings to U.S. offenders.

C. Summary and Conclusions

Empirical evidence fails to support that clinicians can predict violence in either an inpatient or community setting with accuracy equivalent to or exceeding simple actuarial instruments.⁴¹² We agree with William Grove and Paul Meehl's suggestion that "a practitioner

Patients: The HCR-20 Violence Risk Assessment Scheme and the Psychopathy Checklist: Screening Version, 67 J. CONSULTING & CLINICAL PSYCHOL. 917 (1999) (reporting data on 193 patients followed for an average of 626 days, finding that AUCs ranged from .76 for any physical violence to .80 for a violent offense and that patients scoring above the median on the HCR-20 were six times more likely to commit any act of physical violence and thirteen times more likely to commit a violent crime in the community than persons who scored under the median); Tonia Nicholls et al., *Comparing Risk Assessments with Female and Male Psychiatric Outpatients: Utility of the HCR-20 and Psychopathy Checklist: Screening Version*, Paper presented at the Annual Convention of the American Psychological Association (1997).

408. See, e.g., Henrik Belfrage, *Implementing the HCR-20 Scheme for Risk Assessment in a Forensic Psychiatric Hospital: Integrating Research and Clinical Practice*, 9 J. FORENSIC PSYCHIATRY 328 (1998); Mats Dernevik, *Preliminary Findings on Reliability and Validity of the Historical-Clinical-Risk Assessment in a Forensic Psychiatric Setting*, 4 PSYCHOL. CRIME & L. 127 (1998).

409. See, e.g., Henrik Belfrage et al., *Prediction of Violence Within the Correctional System Using the HCR-20 Risk Assessment Scheme*, 11 J. FORENSIC PSYCHIATRY 167 (2000); Kevin S. Douglas & Christopher D. Webster, *The HCR-20 Violence Risk Assessment Scheme: Concurrent Validity in a Sample of Incarcerated Offenders*, 26 CRIM. JUST. & BEHAV. 3 (1999).

410. See, e.g., Mats Dernevik et al., *Violent Behaviour in Forensic Psychiatric Patients: Risk Assessment and Different Risk-Management Levels Using the HCR-20*, 8 PSYCHOL. CRIME & L. 93 (2002).

411. See Borum, *supra* note 346, at 950.

412. But see Thomas R. Litwack, *Actuarial Versus Clinical Assessments of Dangerousness*, 7 PSYCHOL. PUB. POL'Y & L. 409 (2001). Litwack suggests that "[a]lthough . . . actuarial predictions have been demonstrated to be superior to clinical predictions for a fairly wide range of prediction tasks," it is not clear they are superior for dangerousness predictions. *Id.* at 410. "If actuarial instruments . . . are to be used in the fairest and most effective manner, they must be validated in a far more precise manner than has occurred." *Id.* Thus, Litwack concludes that "it is premature to substitute actuarial for clinical assessments of dangerousness." *Id.*

who claims not to need any statistical or experimental studies but relies solely on clinical experience as adequate justification, by that very claim is shown to be a nonscientifically minded person whose professional judgments are not to be trusted.”⁴¹³ The claim that many clinical judgments in violence predictions remain untested⁴¹⁴ is not a compelling argument for their continued use in a variety of legal contexts. Untested assumptions, supported only by guesswork or speculation, are appropriately prohibited under the Federal Rules of Evidence.⁴¹⁵ In the next Part, we will evaluate this assertion by examining the evidentiary admissibility standards prescribed in the Federal Rules of Evidence and *Daubert*.

VIII. EVIDENTIARY ADMISSIBILITY ANALYSIS

The Court in *Daubert* interpreted the Federal Rules to imply a mandate that members of the federal judiciary perform a “preliminary assessment” to determine whether the proffered testimony meets a standard of evidentiary reliability through which scientific evidence must pass.⁴¹⁶ Courts, however, have been reluctant to apply

413. Grove & Meehl, *supra* note 338, at 320.

414. See, e.g., Litwack et al., *supra* note 355, at 247 (suggesting that the “validity of clinical assessments of dangerousness in a variety of important contexts remains untested”).

415. But cf. Thomas R. Litwack, *Communications Regarding Risk*, 52 AM. PSYCHOL. 1245 (1997) (“In any event, the necessity of relying on categorical and unproven (although hopefully informed and rational) clinical judgments regarding many groups of individuals (e.g., many civil committees, or potential civil committees, and insanity acquittees) remains.”). Professional standards have also addressed limits on prediction testimony. See AM. PSYCHOL. ASSOC., ETHICAL PRINCIPLES OF PSYCHOLOGISTS AND CODE OF CONDUCT §§ 2.04(b), 7.04 (Section 2.04(b) states, “Psychologists recognize limits to the certainty with which diagnoses, judgments, or predictions can be made about individuals,” and section 7.04 states, “Whenever necessary to avoid misleading, psychologists acknowledge the limits of their data or conclusions.”); see also Committee on Ethical Guidelines for Forensic Psychologists, *Specialty Guidelines for Forensic Psychologists*, 15 LAW & HUM. BEHAV. 655, 665 (1991) (requiring forensic psychologists to be “aware that their own professional observations, inferences, and conclusions must be distinguished from legal facts, opinions, and conclusions.” They must further be “prepared to explain the relationship between their expert testimony and the legal issues and facts of an instant case.”). The Association for the Treatment of Sexual Abusers (“ATSA”) also issued a cautionary statement. See ASSOC. FOR THE TREATMENT OF SEXUAL ABUSERS, ATSA PRACTITIONER’S HANDBOOK 2, 20 (1993) (instructing its members to “avoid drawing conclusions or rendering opinions that exceed the present level of knowledge in the field or the expertise of the evaluator,” and to “be very cautious in offering predictions of criminal behavior for use in imprisoning or releasing individuals.”). ATSA further requires that prediction testimony include “the acts being predicted; the estimated probability that these acts will occur during a given period of time; and the facts on which these predictive judgments are based.” *Id.*

416. *Daubert v. Merrell Dow Pharms., Inc.*, 509 U.S. 579, 592 (1993). The Court

admissibility criterion to expert psychological testimony.⁴¹⁷ Professor Slobogin hypothesized that the courts readily except psychological testimony because it has not been regarded as “novel.”⁴¹⁸ Recent rulings, however, have rejected the view that evidentiary admissibility standards apply only to novel testimony.⁴¹⁹

A second reason that courts may have been reluctant to apply admissibility criteria to psychological testimony is that the justice system has become dependent on psychologists “to establish the presence or absence of mental disorders and the causal connections between such disorders and criminal or tortious conduct.”⁴²⁰ The sentiment that the courts are unwilling to “disinvent the wheel”⁴²¹ is pervasive and difficult

described the duty of the trial judge as follows:

Faced with a proffer of expert scientific testimony, then, the trial judge must determine at the outset, pursuant to Rule 104(a), whether the expert is proposing to testify to (1) scientific knowledge that (2) will assist the trier of fact to understand or determine a fact in issue. This entails a preliminary assessment of whether the reasoning or methodology underlying the testimony is scientifically valid and of whether that reasoning or methodology properly can be applied to the facts in issue.

Id. at 592–93.

417. One notable exception was *In re Wilson*, 33 CRIM. L. REP. (BNA) 2115 (D.C. Super. Ct. Apr. 14, 1983). In this case, the District of Columbia Superior Court applied the *Frye* “general acceptance” test to dangerous testimony and held that because professional organizations, including the American Psychiatric Association and the American Psychological Association, questioned the reliability of prediction testimony, it should not be admissible in civil commitment proceedings. *Id.*; see *Frye v. United States*, 293 F. 1013 (D.C. Cir. 1923); cf. *Barefoot v. Estelle*, 463 U.S. 880, 901 (1983) (noting that “[n]either petitioner nor the [APA] suggests that psychiatrists are always wrong with respect to future dangerousness, only most of the time” and incredibly upholding the admission of Dr. Grigson’s testimony).

418. See Slobogin, *supra* note 224, at 138 n.155 (“One reason testimony by mental health professionals is accepted so readily by the courts is that it has never been regarded as ‘novel’; it has long been an everyday feature of the justice system.”).

419. See *Gammill v. Jack Williams Chevrolet, Inc.*, 972 S.W.2d 713, 721 (Tex. 1998) (finding that the rules of evidence regarding admission of expert scientific testimony should not apply only to evidence deemed novel or unconventional, but ruling that all scientific expert testimony must meet reliability and relevancy requirements). The court supported this position on several grounds: first, that differentiating between novel and conventional evidence would be a difficult task; second, that it would be illogical to require some expert opinion testimony to be reliable but other not; and third, that the majority of federal courts have applied relevancy and reliability requirements to all expert testimony. *Id.* at 721.

420. See James M. Doyle, *Applying Lawyers’ Expertise to Scientific Experts: Some Thoughts About Trial Court Analysis of the Prejudicial Effects of Admitting and Excluding Expert Scientific Testimony*, 25 WM. & MARY L. REV. 619, 623 (1984) (quoting Comment, *The Psychologist as Expert Witness: Science in the Courtroom?*, 38 MD. L. REV. 539, 544–46 (1979)).

421. See *Barefoot*, 463 U.S. at 896.

to counter. The analysis provided in this Article does not support a wholesale exclusion of predictions of dangerousness; rather, only unreliable or speculative predictions would be deemed inadmissible. This moderate position should make the adoption of admissibility analysis more probable.

A final reason that psychological testimony may have been excused from the admissibility standards is the belief that expert testimony of this kind does not sway the trier of fact as easily as does more complex, less intuitive testimony.⁴²² This argument is not credible in light of the weight of evidence indicating the contrary.⁴²³

Judges have been assigned the gatekeeping role. The arguments for failing to perform admissibility assessments are not compelling. Courts have begun to recognize the importance of applying the standards to all types of testimony, and as this trend continues, experts should expect to demonstrate predictive accuracy or have their testimony excluded. Judges and attorneys must also be prepared to appropriately challenge proffered expert testimony of this kind. In the last Part, we will review a few circumstances in which this testimony would likely be presented and emphasize the admissibility criterion that might be at issue.

IX. STANDARDS OF PROOF

In the second component of this Article, we have expressed a clear preference for prediction testimony informed by actuarial data. In this final Part, we will analyze the most significant prediction context in which dangerousness testimony is frequently offered—capital punishment—and we will discuss the admissibility and standard of proof issues likely to arise when the trier of fact evaluates this type of evidence in a capital sentencing case.

“Future dangerousness” is currently identified as a statutory aggravating factor in capital cases in seven states.⁴²⁴ In all of these

422. See Slobogin, *supra* note 224, at 138 n.155.

423. See, e.g., *Barefoot*, 463 U.S. at 916 (Blackmun, J., dissenting) (describing an “impressionable jury” influenced by the “inevitable untouchability of a medical specialist’s words”); *People v. Murtishaw*, 29 Cal. 3d 733, 774 (1981) (holding that expert predictions of future dangerousness are prejudicial and unduly influence the lay jury).

424. See IDAHO CODE ANN. § 19-2515(9)(h) (2001) (“The defendant, by prior conduct or conduct in the commission of the murder at hand, has exhibited a propensity to commit murder which will probably constitute a continuing threat to society.”); OKLA. STAT. ANN. tit. 21, § 701.12(7) (West 1983) (“The existence of a probability that the defendant would commit criminal acts of violence that would constitute a continuing threat to society.”); OR. REV. STAT. § 163.150(1)(b)(B) (1996) (“Whether there is a probability that the defendant would commit criminal acts of violence that would constitute a continuing threat to society.”);

jurisdictions, the state must demonstrate future dangerousness “beyond a reasonable doubt”⁴²⁵ to secure a death sentence.⁴²⁶ The quantification associated with each burden of proof becomes critical when judges evaluate prediction testimony based on actuarial instruments. As an illustrative vehicle, we will review a hypothetical case using Wyoming’s capital sentencing statute, which requires jurors to be convinced “beyond a reasonable doubt” that a defendant will “likely” commit continued acts of criminal violence.⁴²⁷

TEX. CODE CRIM. PROC. ANN. art. 37.071(2)(b)(1) (West Supp. 2006) (“[W]hether there is a probability that the defendant would commit criminal acts of violence that would constitute a continuing threat to society.”); VA. CODE ANN. § 19.2-264.2(1) (2004) (“[T]here is a probability that the defendant would commit criminal acts of violence that would constitute a continuing serious threat to society.”); WYO. STAT. ANN. § 6-2-102(h)(xi) (2007) (“The defendant poses a substantial and continuing threat of future dangerousness or is likely to commit continued acts of criminal violence.”). In Washington, the future dangerousness factor is considered only after a defendant is convicted of aggravated first-degree murder by a finding of at least one other aggravating factor. See WASH. REV. CODE ANN. § 10.95.020 (West Supp. 2002) (listing the other aggravating factors for aggravated first degree murder); *Id.* § 1095.070(8) (“Whether there is a likelihood that the defendant will pose a danger to others in the future.”).

425. See *Addington v. Texas*, 441 U.S. 418, 423 (1979) (addressing the standards of proof and opining, “The standard serves to allocate the risk of error between the litigants and to indicate the relative importance attached to the ultimate decision.”). Attempts to quantify the burden of proof have resulted in some general parameters (for example, beyond a reasonable doubt = 95% chance of future violence; clear and convincing evidence = 75% chance of future violence; and preponderance of the evidence = 51% of the evidence). See C.M.A. McCauliff, *Burdens of Proof: Degrees of Belief, Quanta of Evidence, or Constitutional Guarantees?*, 35 VAND. L. REV. 1293, 1322, 1327 (1982) (discussing the differences between “beyond a reasonable doubt” at 95% versus 99% certainty). This author went on to survey members of the judiciary to determine their understanding of the probability attached to each designated burden of proof. Judicial consensus appears to be that “reasonable doubt” rests somewhere between 80% to 100%. *Id.* at 1324–27.

426. Several authors have considered the ethics of participation in capital sentencing hearings. See, e.g., *Dix*, *supra* note 27 (concluding that expert predictions of dangerousness in capital sentencing hearings violate evidentiary and constitutional standards); *Barefoot*, 463 U.S. at 916 (Blackman, J., dissenting).

In the present state of psychiatric knowledge, this is too much for me. One may accept this in a routine lawsuit for money damages, but when a person’s life is at stake[,] . . . a requirement of greater reliability should prevail. In a capital case, the specious testimony of a psychiatrist, colored in the eyes of an impressionable jury by the inevitable untouchability of a medical specialist’s words, equates with death itself.

Barefoot, 463 U.S. at 916 (Blackman, J., dissenting)

427. See WYO. STAT. ANN. § 6-2-102(h)(xi). For the purposes of this exercise, we will assign “likely” to 70% probability. As the specificity of scientific evidence increases, legislators should be encouraged to be more specific in statutes incorporating dangerousness as a criterion. These statutes should include the type of harm that is being predicted

Suppose that a psychologist proffers testimony in a capital sentencing hearing concerning the defendant's risk of future violence based on his score of the Violence Risk Assessment Guide.⁴²⁸ She opines that "[t]he defendant's score places him in the sixth category, or third from the highest risk group. Among offenders in prior studies, approximately 42% reoffended violently within an average of 10 years following release." The probability estimation generated by quantifying the burden of proof standards indicates that, in this example, the state fails to meet its burden of proof in the case.⁴²⁹

Sexual offender civil detention statutes also generally rely on the "beyond a reasonable doubt" standard. One of the most significant questions related to these predictions involves the specificity of the prediction task. Typically, sexual, rather than general, violence is being predicted; this should be reflected in the expert's assessment.⁴³⁰

(physical, emotional, property); the degree of probability that this behavior will occur (50%, 95%, or more likely than not) and the period of time that risk is predicted to occur (a year, indefinite). See Slobogin, *supra* note 224, at 104 n.28.

428. See QUINSEY ET AL., *supra* note 33.

429. That is, an actuarial prediction of .42 would fail to meet the statutory standards that "beyond a reasonable doubt" (95%) the defendant would "likely commit a violent act" (70%) (.95 x .70 = .665; .665 > .42), thus, the aggravating circumstance would not be met. See Slobogin, *supra* note 224, at 104 n.28.

This case presumes that the judge had previously conducted an admissibility assessment on the proffered testimony. In this example, the judge must have previously assessed the ability of the expert's opinion to assist the trier of fact and would have determined that the expert demonstrates some expertise in both prediction of violence and actuarial methods. At first glance, relevancy and fit appear to be fairly obvious in this case; however, recent research examining the use of actuarial instruments to predict institutional violence in correctional settings suggests that predictive accuracy is significantly lower than originally expected and has led one group of researchers to contend that, at least in the context of predicting violence by capital defendants, the VRAG, PCL-R, and HCR-20 have resulted in judgments that are "highly inaccurate" and "gross overestimations of risk." See John F. Edens et al., *Predictions of Future Dangerousness in Capital Murder Trials: Is It Time to "Disinvent the Wheel?"*, 29 LAW & HUM. BEHAV. 55 (2005). If these obstacles to admissibility were satisfactorily addressed, reliability questions would then be significant. The expert must demonstrate that he used appropriate methodology, that he has a good foundation for the use of this instrument, and that the use of the instrument is logically connected to the defendant. Jurists should review the original empirical studies that the expert relied upon as justification for using this instrument. Differences between characteristics of the defendant (e.g., race, age, severity of the current offense, length of time since publication of the studies) and environment (e.g., community versus high security correctional institution) should be evaluated to determine whether the connective reliability criterion has been met. Judges should ask about the type of information relied upon in generating this score, paying close attention to the reliability of the underlying data. Did the expert verify his sources whenever possible or did he rely on hearsay or other inadmissible evidence?

430. Several new actuarial instruments have recently been introduced to predict future sexual violence, such as the Minnesota Sex Offender Screening Tool-Revised ("MnSOST-R")

The lowest threshold for predictions of violence involves the commitment of individuals found “not guilty by reason of insanity.” A unique aspect of these cases is the shift of the burden of proof from the state to the defendant. The broad definition of violence established by the courts in these cases⁴³¹ again should be reflected in the assessment.⁴³²

This hypothetical case and accompanying summaries illustrate the importance of specificity in sentencing statutes.⁴³³ Predictions of dangerousness based on actuarial data preclude attempts by the expert to usurp the decision maker’s role. The degree of accuracy necessary to justify a particular judicial action (civil commitment, maximum sentencing, or even death) is a value judgment best made by elected or appointed officials, not by an expert whose implicit values are cloaked in the guise of medical certainty.

X. CONCLUSION

Members of the judiciary face a difficult but important task when evaluating proffered expert testimony to determine whether it meets admissibility standards. Legislators and the courts have recognized the importance of assuring that proffered expert testimony meets minimum reliability standards. But what our argument attempts to add is recognition of the importance of perspective and contextual complexity in assuring that reliability. Nothing in this argument diminishes the importance of experts; courts very much need the advice of specialized opinions. But those opinions often have an aura to outsiders, regardless

and the Sexual Offender Risk Appraisal Guide (“SORAG”). See QUINSEY ET AL., *supra* note 33, at 119–36. Judges should ascertain the characteristics of the validation sample and the similarities or differences between study participants and the individual currently being evaluated. It may be helpful to hear the expert describe and explain discrepancies in risk assessment on each of the two instruments. See Cornwell, *supra* note 385 (discussing the problems with dangerousness prediction in this context). See generally Steven I. Friedland, *On Treatment, Punishment, and the Civil Commitment of Sex Offenders*, 70 U. COLO. L. REV. 73 (1999) (questioning the use of violence prediction as a means of social control).

431. See *Jones v. United States*, 463 U.S. 354, 364–65 (1983) (“We do not agree with the petitioner’s suggestion that the requisite dangerousness is not established by proof that a person committed a non-violent crime against property.”).

432. Judges should pay special attention to the appropriateness of using an actuarial instrument in these cases. Many of the actuarial data have been collected from individuals with a history of *serious* violence. The extent to which the defendant differs in this way should raise connective reliability concerns.

433. Quantification of this kind may result in statutes being deemed unconstitutional on the grounds that they are unacceptably vague. See, e.g., *Williamson v. Reynolds*, 904 F. Supp. 1529, 1569–71 (E.D. Okla. 1995) (granting writ of habeas corpus in a capital case in which the statutory aggravating circumstances of future dangerousness, without guidance on the definition of the language of the statute, was unconstitutionally vague).

of the acuity of those non-experts in their own realms, which exaggerates the authority of the expertise.

We have argued here that while, in many instances, prediction testimony based on clinical judgment should be deemed inadmissible,⁴³⁴ prediction testimony based on actuarial data can assist the court and meet admissibility criterion under most circumstances. One caveat to this admission is that opposing expert testimony must be presented to prevent the expert's opinions from being unfairly prejudicial and, thus, inadmissible.

In the context of reviewing admissibility criterion, members of the judiciary should conduct a preliminary assessment to ascertain whether the proffered expert testimony meets the criterion described in the Federal Rules of Evidence and subsequent case law. Judges should examine the qualifications of a witness to determine whether he or she has special expertise in predicting dangerousness; they should determine whether the testimony is based on clinical or actuarial data, and they should review the studies and literature that the expert contends support his opinion.⁴³⁵ The size of the construction and validation samples, number of independent validation studies, and degree of similarity between the study participants and the individual presently evaluated should all be considered in determining the reliability of the opinion. Undeniably, these assessments present a daunting task to trial court judges.⁴³⁶ In spite of the difficulties, members of the judiciary must assume this role to protect the reliability and integrity of judicial decisions.

434. Some have raised the question of whether prediction testimony based on clinical judgment has been fairly tested. We would agree that methodological problems inherent in violence prediction studies have limited the generalizability of the research. We concede that in some instances intuitive clinical judgment will beat actuarial prediction; however, we do not believe that clinical judges will ever outperform actuarial models. We based this belief on a very common sense notion of the complexity of decision making and the prevalence of heuristics. Clinical skill is indispensable in violence prediction, but we conclude that actuarial combination of that data will be more accurate, and fairer, than any other method.

435. See Christopher D. Webster & Natalie H. Polvi, *Challenging Assessments of Dangerousness and Risk*, in *COPING WITH PSYCHIATRIC AND PSYCHOLOGICAL TESTIMONY* 1371, 1381 (Jay Ziskin et al. eds., 1995) (providing an excellent example of the kinds of questions that judges or attorneys should ask of expert witnesses regarding the prediction of dangerousness).

436. See Brown, *supra* note 28, at 784 ("Federal judges ruling on the admissibility of expert scientific testimony face a far more complex and daunting task in a post-*Daubert* world than before.") (quoting *Daubert v. Merrell Dow Pharm., Inc.*, 43 F.3d 311, 315 (9th Cir. 1995); see also Ronda R. Harrison, *Daubert and the Federal Judiciary: Understanding Science Methodology* (1998) (unpublished M.A. thesis, Bowling Green State University) (on file with author).

While we appreciate the workload and skill of the judiciary, we still contend that every opportunity should be seized to benefit from hearing the arguments of experts from different schools of thought in a specialty area. Even if no other benefit were achieved from such an effort, the resulting heightened awareness of the diversity of expertise *within* a domain of expertise would give pause to fact finders who otherwise might be tempted to permit the expert to determine the court's judgment. Courts that show their concern for the reliability of experts by making every effort to hear as many different scientific perspectives within a realm of expertise as is practical will be putting experts in their proper epistemological place in our judicial system.

Copyright of *Marquette Law Review* is the property of *Marquette University Law School* and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.